

GI-19745307 264 ---- **HW** **KEKPIIYFKLYRQLPGEKEVAVDDAELK** **PINS**
 ORF84 **WO 2006/078318** **HW** **KEKPIIYFKLYRQLPGEKEVAVDDAELK** **PINSEGO** PCT/US2005/027239
 GI-28810263 264 **HW** **KEKPIIYFKLYRQLPGEKEVAVDDAELK** **PINSEGO**
 GI-21909640 150 **HW** **KEKPIIYFKLYRQLPGEKEVAVDDAELK** **PINSEGO**
 GI-19224141 601 **PHSS** **VEANKEVTIANHRETLTFS** **CKRIWENR** **DRDORPAKIQVQLLQNGQ** **KMPNQIQEV**

GI-19745307 305 --- **Q** **EISVTW** **TNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 ORF84 305 --- **Q** **EISVTW** **TNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 GI-28810263 305 --- **Q** **EISVTW** **TNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 GI-21909640 191 --- **Q** **EISVTW** **TNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 GI-19224141 661 **TKDN** **WSYHEKDL** **PKYDA** **ANQ** **EYKYSV** **EEV** **NVPDGYKVS** **YLGNDI** **FNTRETEFV** **EQNNF**

GI-19745307 349 **GLT** **VTNTYVKPTSG** **HYDIEVTFC** **NGHIDITEDTTPDIVSGENQMK**
 ORF84 349 **GLT** **VTNTYVKPTSG** **HYDIEVTFC** **NGHIDITEDTTPDIVSGENQMK**
 GI-28810263 349 **GLT** **VTNTYVKPTSG** **HYDIEVTFC** **NGHIDITEDTTPDIVSGENQMK**
 GI-21909640 235 **GLT** **VTNTYVKPTSG** **HYDIEVTFC** **NGHIDITEDTTPDIVSGENQMK**
 GI-19224141 721 **NLEFG** **NAETNG** **SGSKI** **IDEDTLT** **SFKG** **KKIWKNDTAB** **NR** **PQAIQV** **QVLYAD** **GVAVEG** **QTK**

GI-19745307 394 **Q** **IEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 ORF84 394 **Q** **IEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-28810263 394 **Q** **IEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-21909640 280 **Q** **IEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-19224141 781 **FIS** **SG** **NEWS** **SFEF** **KNLKKY** **NGTGN** **DIIS** **VKEVT** **VPTGYD** **VTYS** **ANDI** **INTK** **REVI** **TOQG**

GI-19745307 424 --- **DGT** **NSNKYE** **EVEDSR** **PVDTL** **SGLS** **SEQQ** **SGDM** **TEEDS** **SATHIKFS** **KRD**
 ORF84 424 --- **DGT** **NSNKYE** **EVEDSR** **PVDTL** **SGLS** **SEQQ** **SGDM** **TEEDS** **SATHIKFS** **KRD**
 GI-28810263 424 --- **DGT** **NSNKYE** **EVEDSR** **PVDTL** **SGLS** **SEQQ** **SGDM** **TEEDS** **SATHIKFS** **KRD**
 GI-21909640 310 --- **DGT** **NSNKYE** **EVEDSR** **PVDTL** **SGLS** **SEQQ** **SGDM** **TEEDS** **SATHIKFS** **KRD**
 GI-19224141 841 **PKLEI** **EETLPL** **ESGAS** **GGTTT** **VEDSR** **PVDTL** **SGLS** **SEQQ** **SGDM** **TEEDS** **SATHIKFS** **KRD**

GI-19745307 473 **IDG** **KELAGAT** **MELRDS** **SGNTISTWIS** **DGQV** **KDFY** **LMFG** **KYTF** **VETA** **APDC** **YEI** **ATAITFT**
 ORF84 473 **IDG** **KELAGAT** **MELRDS** **SGNTISTWIS** **DGQV** **KDFY** **LMFG** **KYTF** **VETA** **APDC** **YEI** **ATAITFT**
 GI-28810263 473 **IDG** **KELAGAT** **MELRDS** **SGNTISTWIS** **DGQV** **KDFY** **LMFG** **KYTF** **VETA** **APDC** **YEI** **ATAITFT**
 GI-21909640 359 **IDG** **KELAGAT** **MELRDS** **SGNTISTWIS** **DGQV** **KDFY** **LMFG** **KYTF** **VETA** **APDC** **YEI** **ATAITFT**
 GI-19224141 901 **IDG** **KELAGAT** **MELRDS** **SGNTISTWIS** **DGQV** **KDFY** **LMFG** **KYTF** **VETA** **APDC** **YEI** **ATAITFT**

GI-19745307 533 **VNE** **QGOVT** **VNGKAT** **KGDA** **HIVM** **DAYK** **PTK** **SGQ** **VIDIE** **EKL** **PDEQ** **CHSG** **STTEI** **EDSKS**
 ORF84 533 **VNE** **QGOVT** **VNGKAT** **KGDA** **HIVM** **DAYK** **PTK** **SGQ** **VIDIE** **EKL** **PDEQ** **CHSG** **STTEI** **EDSKS**
 GI-28810263 533 **VNE** **QGOVT** **VNGKAT** **KGDA** **HIVM** **DAYK** **PTK** **SGQ** **VIDIE** **EKL** **PDEQ** **CHSG** **STTEI** **EDSKS**
 GI-21909640 419 **VNE** **QGOVT** **VNGKAT** **KGDA** **HIVM** **DAYK** **PTK** **SGQ** **VIDIE** **EKL** **PDEQ** **CHSG** **STTEI** **EDSKS**
 GI-19224141 961 **VNE** **QGOVT** **VNGKAT** **KGDA** **HIVM** **DAYK** **PTK** **SGQ** **VIDIE** **EKL** **PDEQ** **CHSG** **STTEI** **EDSKS**

GI-19745307 593 **SDV** **IIGGQ** **SGSTTEI** **EDSKSS** **SDV** **IIGGQ** **GEVVD** **TTEDT** **QSGMT** **GHS**
 ORF84 593 **SDV** **IIGGQ** **SGSTTEI** **EDSKSS** **SDV** **IIGGQ** **GEVVD** **TTEDT** **QSGMT** **GHS**
 GI-28810263 593 **SDV** **IIGGQ** **GEVVD** **TTEDT** **QSGMT** **GHS**
 GI-21909640 479 **SDV** **IIGGQ** **GEVVD** **TTEDT** **QSGMT** **GHS**
 GI-19224141 1021 **SD** **LIIGGQ** **GEVVD** **TTEDT** **QSGMT** **GHS**

GI-19745307 602 --- **Q** **IVETTE** **DTQTGM** **HGDS** **SGCKTE** **VEDTK** **LVQS** **PHFD** **DNK**
 ORF84 602 --- **Q** **IVETTE** **DTQTGM** **HGDS** **SGCKTE** **VEDTK** **LVQS** **PHFD** **DNK**
 GI-28810263 619 --- **G** **STTKI** **EDSKSS** **SDV** **IIGGQ** **QIVETTE** **DTQTGM** **HGDS** **SGCKTE** **VEDTK** **LVQS** **PHFD** **DNK**
 GI-21909640 539 **GHS** **GSTTKI** **EDSKSS** **SDV** **IIGGQ** **QIVETTE** **DTQTGM** **HGDS** **SGCKTE** **VEDTK** **LVQS** **PHFD** **DNK**
 GI-19224141 1047 --- **G** **STTEI** **EDSKSS** **SDV** **IIGGQ** **QIVETTE** **DTQTGM** **HGDS** **SGCKTE** **VEDTK** **LVQS** **PHFD** **DNK**

GI-19745307 639 **E** **ES** **NS** **EI** **PK** **DK** **PK** **SNT** **SL** **PAT** **GER** **QH** **NM** **FF** **M** **V** **T** **S** **C** **S** **L** **I** **S** **S** **V** **F** **V** **I** **S** **L** **K** **T** **K** **R** **L** **S** **S** **C**
 ORF84 639 **E** **ES** **NS** **EI** **PK** **DK** **PK** **SNT** **SL** **PAT** **GER** **QH** **NM** **FF** **M** **V** **T** **S** **C** **S** **L** **I** **S** **S** **V** **F** **V** **I** **S** **L** **K** **T** **K** **R** **L** **S** **S** **C**
 GI-28810263 676 **E** **ES** **NS** **EI** **PK** **DK** **PK** **SNT** **SL** **PAT** **GER** **QH** **NM** **FF** **M** **V** **T** **S** **C** **S** **L** **I** **S** **S** **V** **F** **V** **I** **S** **L** **K** **T** **K** **R** **L** **S** **S** **C**
 GI-21909640 599 **E** **ES** **NS** **EI** **PK** **DK** **PK** **SNT** **SL** **PAT** **GER** **QH** **NM** **FF** **M** **V** **T** **S** **C** **S** **L** **I** **S** **S** **V** **F** **V** **I** **S** **L** **K** **T** **K** **R** **L** **S** **S** **C**
 GI-19224141 1104 **E** **ES** **NS** **EI** **PK** **DK** **PK** **SNT** **SL** **PAT** **GER** **QH** **NM** **FF** **M** **V** **T** **S** **C** **S** **L** **I** **S** **S** **V** **F** **V** **I** **S** **L** **K** **T** **K** **R** **L** **S** **S** **C**

FIGURE 54A

172/487

GI-19224WO 2006/078318-MKKNKLLLLATAILATALGTASLNQNVKAETAGVVSSCOLTKKRSITPCT/US2005/027239

ORF80 1 MKKNKLLLLATAILATALGTASLNQNVKAETAGVVICKSLQVTKMT-YDDDEVLM
 GI-21909636 1 MKKNKLLLLATAILATALGTASLNQNVKAETAGVSENKLVKNTFDSYTDNEVLM
 GI-28810259 1 MEREKMKKNKLLLLATAILATALGTASLNQNVKAETAGVSENKLVKNTFDSYTDNEVLM
 GI-19745303 1 MKKNKLLLLATAILATALGTASLNQNVKAETAGVSDGSLVKKTFPSYTDKVLN
 GI-13621428 1 MKLRHLLLTGAALTSFAAT-VHGETVVNCAKLTVTKNLDELVNSN-ALN

GI-19224137 56 PNTDYTFSSVNPDSAACTESN-LPIKPGIAVN-NQDIKVSYSNTDRTSGNEHQVVVDFMK
 ORF80 60 PETAFETFEIPDMTASGREGS-LDIRNGIVEGLDKQVTVKYKNTDRTSQNTNLAQDFDFSK
 GI-21909636 56 PNADYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTNTDRPDSKVKSTEFDFSK
 GI-28810259 61 PNADYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTNTDRPDSKVKSTEFDFSK
 GI-19745303 56 PNADYTFKVEADSTAKGRTKDGLEIKPGIVNGLT-EQILSYTNTDRPDSKVKSTEFDFSK
 GI-13621428 48 PNTDFTFEIPDTTVNEDGNKFKG-VALNTPMTKVITYTNSDNGGSNTKTAEFDFSE

GI-19224137 114 VTFFPSVGTYRYVVTENKGTAE-CVTDYDTKILVDVYVGN-NEKGGLEPKYIVSNKCD SA
 ORF80 119 VKFPAIGVYRYMVSEKNDKED-3ITYDDKRI/TVDVYVGNKANNEDGFEVL YIVSKEGTSS
 GI-21909636 115 VVFPFGIGVYRYTVSEKQGDVE-3ITYDTRKI/TVDVYVGN-KEGGGFEPKEIVSKEQGT
 GI-28810259 120 VVFPFGIGVYRYTVSEKQGDVE-3ITYDTRKI/TVDVYVGN-KEGGGFEPKEIVSKEQGT
 GI-19745303 116 VKFPGVGVYRYTVSEVNCNNA-3IAYDSQQ/TVDVYVGN-REDGGFEAKYIVSTEGGQS
 GI-13621428 103 VTFEKPGVYRYKVTEEKIDKVP-3VSYDTSYTVQVHVLWN-BEQOKPVATYIVGYKEGS-

GI-19224137 171 TNEPTOFNNSFETTSKRTENBVTGNTGDKHAFITFTLTLOPNEYWEASSVVKIEENGQ--
 ORF80 178 TKNPIEFINSIKTTSLKTEKQITGNAQDRKESFNFTLTLOPSEYKRTGSVVKIEQDCS--
 GI-21909636 172 VKNPVMFNNSFATTSKLVKINVSQNTGELQKEFDFTLTLESTNEKNDQIVSLQKCN--
 GI-28810259 177 VKNPVMFNNSFATTSKLVKINVSQNTGELQKEFDFTLTLESTNEKNDQIVSLQKCN--
 GI-19745303 173 DKKPVLEKNEFDTTSLKVTNKVTGNTGHEHRSFSFTLTLESTNECEBERGOVNILOCGE--
 GI-13621428 161 -KVPFOFKNSIDSTTLTVKHKVSCITGDRSDENEGTLTKANQYKASEKVMIERTTKGG

GI-19224137 229 ----TNDVKICEAYKFTLNDSSSVILSKLPVGINVKVDEAEANQCGYTTATLNDG--EK
 ORF80 236 ----KNDVILGTPKFTLGHCKSVMLSKLPIGINVLESEDEANKDGYTTATLNEQKKEK
 GI-21909636 230 ----KDEVKIGTPYKFKLNGESTOLDKLPVGEIYKVNEMAEANKDGYKTASLNEC-DGQ
 GI-28810259 235 ----KDEVKIGTPYKFKLNGESTOLDKLPVGEIYKVNEMAEANKDGYKTASLNEC-DGQ
 GI-19745303 231 ----TKKVICETLFTLNDKESVILSOLPVGEIYKVNTEEDVTHDGYKTSATLNDG-D-V
 GI-13621428 220 QAPVQTEASTDQTHFTLNDGESIKVTNLPVGVDEIVTEDDYKSEKNTINVEVSPQDGA

GI-19224137 283 LSTYNLC-QEHTDKTAEIVVTNNRDTQVPTGVVGTLPFAVL SIVAIGGVYITRKK
 ORF80 292 SSBFTLSTONKRTDESAD EIVVTNNRDTQVPTGVVGTLPFAVL SIVAIGGVYITRKK
 GI-21909636 285 SKMYOLD-MEQRTDESAD EIVVTNNRDTQVPTGVVGTLPFAVL SIVAIGGVYITRKK
 GI-28810259 290 SKMYOLD-MEQRTDESAD EIVVTNNRDTQVPTGVVGTLPFAVL SIVAIGGVYITRKK
 GI-19745303 285 TDCYNLC-DSKRTDSTDEIVVTNNRDTQVPTGVVGTLPFAVL SIVAIGGVYITRKK
 GI-13621428 280 KNIAGNSTEQETSDKDMILTEINNKDEVPPTGVAMTVAPYIALCIVAVGGALYEVNKKN

GI-19224137 342 A
 ORF80 352 A
 GI-21909636 344 A
 GI-28810259 349 A
 GI-19745303 344 A
 GI-13621428 340 A

FIGURE 55

173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLF SVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLF SVVMILTMLAFNOTVLAKDSTV
ORF82	1	LLFQRVKIFLLTIVLSLSVLFKNNERRRLRKYWK
GI-19745305	1	MRKYWK MLF SVVMILTMLAFNOTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFSALESIDAMKTIEEITTAGSGKASFSPLTFTTVGQYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFSALESIDAMKTIEEITTAGSGKASFSPLTFTTVGQYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFSALESIDAMKTIEEITTAGSGKASFSPLTFTTVGQYTYRV
ORF82	61	QTSISVENVLERAGDSTPFSALESIDAMKTIDEITTAGSGKASFSPLTFTTVGQYTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFSALESIDAMKTIDEITTAGSGKASFSPLTFTTVGQYTYRV

GI-21909638	80	YQKPSQNKDYQADTTVFDVLVYVYTYDEDCTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-28810261	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDCTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19224139	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDCTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
ORF82	121	YQKPSQNKDYQADTTVFDVLVYVYTYDEDCTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19745305	92	YQKPSQNKDYQADTTVFDVLVYVYTYDEDCTLVAKVISRRAGDEEKSAITFKPKRLVKPIP

GI-21909638	140	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-28810261	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19224139	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
ORF82	181	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19745305	152	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs	exp
	(=)	(*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:==*==
56	4	3:==*==
58	5	2:==*==
60	3	2:==*==
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:==
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:==
98	0	0:

FIGURE 57

PCT/US2005/027239

100	0	0:
102	0	0:
104	0	0:
106	1	0:=
108	0	0:
110	0	0:
112	0	0:
114	0	0:
116	0	0:
118	0	0:
>120	1	0:=

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

init1 initn opt z-sc E(54)...

```

/home/morana/gas/pili/align/gi-50913505.pep      Begin: 1  End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 3452.1 9.6e-189
/home/morana/gas/pili/align/gi-19224141.pep      Begin: 48  End: 144
! gi|19224141|gb|AAL86412.1|AF447492_... 63 100 159 105.9 0.023
/home/morana/gas/pili/align/gi-21909640.pep      Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morana/gas/pili/align/gi-13621428.pep      Begin: 57  End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morana/gas/pili/align/gi-50913506.pep      Begin: 33  End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morana/gas/pili/align/gi-13621432.pep      Begin: 14  End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morana/gas/pili/align/gi-19745301.pep      Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morana/gas/pili/align/gas15.pep           Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morana/gas/pili/align/gi-21909636.pep      Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morana/gas/pili/align/gi-28810259.pep      Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morana/gas/pili/align/gi-19224139.pep      Begin: 90  End: 143
! gi|19224139|gb|AAL86410.1|AF447492_... 43 43 54 58.9 8.9
/home/morana/gas/pili/align/gi-19745305.pep      Begin: 96  End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morana/gas/pili/align/orf82.pep           Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morana/gas/pili/align/gi-21909638.pep      Begin: 84  End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List

```

gi-50913505.pep

/home/morana/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189

>>/home/morana/gas/pili/align/gi-50913505.pep (1036 aa)

initn: 6697 init1: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189

Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap

(1-1036:1-1036)

10 20 30 40 50 60
gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLPLIRRLGLQMIQTQTKVIPQEIIVT

FIGURE 57A

FIGURE 57B

PCT/US2005/027239

```

      670      680      690      700      710      720
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINS DTASSSPVVLKYLSGEEHYYGITDTAELE
|||||
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINS DTASSSPVVLKYLSGEEHYYGITDTAELE
      670      680      690      700      710      720

      730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
|||||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
      730      740      750      760      770      780

      790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFSKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
|||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFSKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
      790      800      810      820      830      840

      850      860      870      880      890      900
gi-50913505. EMGDSDDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
|||||
gi-50913505. EMGDSDDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
      850      860      870      880      890      900

      910      920      930      940      950      960
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
|||||
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
      910      920      930      940      950      960

      970      980      990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
|||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
      970      980      990      1000      1010      1020

      1030
gi-50913505. TVALLFYRRQHRKKQY
|||||
gi-50913505. TVALLFYRRQHRKKQY
      1030

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

```

SCORES: Init1: 63   Initn: 100   Opt: 159   z-score: 105.9 E(): 0.023
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 100 init1: 63 opt: 159 Z-score: 105.9 expect(): 0.023
Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
(895-990:48-144)

```

```

      870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKK-IV
      :| :| :| :| :| :|
gi-19224141. FILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGGKTS
      20      30      40      50      60      70

      930      940      950      960      970      980

```

FIGURE 57C

```

      1000      1010      1020      1030
gi-50913505. KIQQLPSSGGRGSQIFLIIVGSMATATVALLFYRROHRRKKOY

```

FIGURE 57D

PCT/US05/27239

179/487

|: |:|:

gi-21909640. DAYK-PTKGSQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGM
450 460 470 480 490 500

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 33 Initn: 33 Opt: 91 z-score: 75.6 E(): 1.1
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 33 init1: 33 opt: 91 Z-score: 75.6 expect(): 1.1
Smith-Waterman score: 95; 19.9% identity in 271 aa overlap
(568-819:57-318)

540 550 560 570 580 590
gi-50913505. NKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTAL
gi-13621428. VNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALNTPMTKVITYTNS
30 40 50 60 70 80

600 610 620 630 640 650
gi-50913505. HNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYR-----SGNGS
gi-13621428. DKGGSNTKTAEFDFSEVTFEKPVGYYKVVTEEKIDKVPVSYDTTSYTVQVHVLWNEEQQ
90 100 110 120 130 140

660 670 680 690 700
gi-50913505. SNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINS DTAS-SSPVVLKYLSGEEH
gi-13621428. KPVATYIVGYKEGSKVPI-QFKN---SLDSTTLTVKKKVS GTGGDRSKDFNFGTLTKANQ
150 160 170 180 190 200

710 720 730 740 750
gi-50913505. YYGITDTAELEKTLN---KIVEDSKLSQL---GISDSL SQYVDYYDKQPDVLVT---R
gi-13621428. YYKASEKVMIEKTTKGGQAPVQTEASIDQLYHFTLKDGESIKVTNLPVGVVDYVVTEDDYK
210 220 230 240 250 260

760 770 780 790 800 810
gi-50913505. KSKVNDETEILYQKDQVQEAGKDIIDKVVFPTPKTTSQPKGKVTLTFSKYVDDEYTYTL
gi-13621428. SEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMT-----ITFTNKKDFEVPTGVAMTV
270 280 290 300 310

820 830 840 850 860 870
gi-50913505. SENVKASDEAYEKYKDN EGRYSEM GDS DTDYGTNQTSSGKGG LPSNSDASVNYMADGREQ
gi-13621428. APYIALGIVAVGGALYFVKKKNA
320 330 340

gi-50913505.pep

/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
MGAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 71.3 E(): 1.9
>>/home/morama/gas/pili/align/gi-50913506.pep (556 aa)

FIGURE 57E

initn: 149 initl: 70 opt: 86 Z-score: 71.3 expect(): 1.9
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(503-966:33-428)

```

      480      490      500      510      520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRRD DAVKNSLLGVNG-----L
                                     |:: || :: :||::: :: ::::|
gi-50913506. NRRETVREKILITAKKIMLACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
      10      20      30      40      50      60

      530      540      550      560      570      580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNRDAELLKGWSTNS
      :  ::  |  |  |::  |  |  :  ::  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. YKIGEGVYNTNGDSFINFK-----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
      70      80      90     100     110

      590      600      610      620      630      640
gi-50913505. LLDENTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
      :  |  ::  :  |::  :  |||  :  |||  |::  :  |  ||  :  :
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS-----YNGEGLNLTVK
      120     130     140     150     160

      650      660      670      680      690      700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDFKARYPNLSIYSLGVSKDINS DTASSSPVVLKYLSGE
      |  :|  ::|  :|  ::::|  |:  |:::  |:  |::|  |:|  :|  :|  :|  :|
gi-50913506. NIDS--KSNLYLGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
      170     180     190     200     210

      710      720      730      740      750      760
gi-50913505. EHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDQKQPDVLVTRKSKVND ETE
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. LPSY-----TKEAVNKT VY-----VSDNMSEGLTF--NFNSLTVEWKGMANITE
      220     230     240     250     260

      770      780      790      800      810      820
gi-50913505. ILYQKDQVQEAGKDIIDKVVF TPKTTSQPKGKVTLTFKSDYKVDDEYTYT LSFNVKASDE
      |  |:::|  :|  |::  |:  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      270     280     290

      830      840      850      860      870      880
gi-50913505. AYEKYKDNEGRYSEMGSDDTDYGTNQTSSGKGCLPSNSDASVNYMADGREQKLPYKH PVI
      :|:  :|::  :|  |:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. SYKAVVNNKAI VGEENPNKAEEFFYSNNPTKCN TYDNLDKKPKD-KNGITSKEDSK----
      300     310     320     330     340     350

      890      900      910      920      930      940
gi-50913505. QVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGT TGSNGQLNFKYLQKGKTY Y
      |  |  |::|  |||:  ::  |  |:  |:  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. IVYTYQIAERKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAISTQVSSGK-YK
      360     370     380     390     400

      950      960      970      980      990     1000
gi-50913505. LYETKAKLGYTL PENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
      :  |  ||  ||:|  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
      410     420     430     440     450     460
```

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene

FIGURE 57F

51

FIGURE 57G

PCT/US2005/027239

440

450

460

470

480

490

1020

1030

gi-50913505. VGSMTATVALLFYRRQHRKKQY

gi-19745301. GLTTDGAIYLWLLLLLVLLGLWVWLIGRKGLKND

500

510

520

gi-50913505.pep

/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 43 Initn: 68 Opt: 69 z-score: 61.4 E(): 6.6
>>/home/morama/gas/pili/align/gas15.pep (762 aa)
initn: 68 init1: 43 opt: 69 z-score: 61.4 expect(): 6.6
Smith-Waterman score: 100; 21.4% identity in 252 aa overlap
(641-873:492-739)

	620	630	640	650	660	
gi-50913505.	AKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI					
gas15.pep	HIAGRDLFKYTVKPRDTPDPTFLKHIKKVIEKGYRE-KGQAIEYSGLTETQLRAATQLAI					
	470	480	490	500	510	520
	670	680	690	700	710	720
gi-50913505.	DEF--KARYPNLSIYSLGVSKDINSSTASSPVVLKYLSGEEHYGTDTAELKTLNKI					
gas15.pep	YYFTDSAELDKDKLKDYGFGDMNDSTLAVAKILVEY-AQDSNPPQLTDLDFIPNNNKY					
	530	540	550	560	570	
	730	740	750	760	770	
gi-50913505.	VEDSKLSQLGISDSLSQYVDYDQKQPDVLVT----RKSKVN---DETEILYQKDQVQEA					
gas15.pep	--QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN					
	580	590	600	610	620	630
	780	790	800	810	820	830
gi-50913505.	GKDIIDRVVFTPKTTSQPK-GKVTLTFSKSDYKVDDE-YTYTSLFNVKASDEAYEKYKDNE					
gas15.pep	KQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLOGLPEGYSYLVKETDSEGYKVKVNS					
	640	650	660	670	680	690
	840	850	860	870	880	890
gi-50913505.	GRYSEMGSDDTDYGTNQT----SSGKGLPSNSDASVN-YMADGREQKLPYKHPVIQVKT					
gas15.pep	QEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGIWGIHTIR					
	700	710	720	730	740	750
	900	910	920	930	940	950
gi-50913505.	VPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLOKGRKTYLYET					
gas15.pep	IRKHD					
	760					

gi-50913505.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

FIGURE 57H

SCORES Initl: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1
>>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
initn: 31 initl: 31 opt: 62 Z-score: 60.8 expect(): 7.1
Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
(181-305:176-298)

	160	170	180	190	200	210
gi-50913505.	DITTTLTFFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR					
				:: :	::: :	:
gi-21909636.	VDVYVGNKEGGGFEPKFIVSKEQGTVDKPKPVNFNNSFATTSLKVKKNVSGN-----TGE					
	150	160	170	180	190	
	220	230	240	250	260	
gi-50913505.	LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL					
	:	: : :	: ::	: :: :	: : :	: :
gi-21909636.	LQKEFDFTLTNLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNGESIQLDKLPVGI					
	200	210	220	230	240	250
	270	280	290	300	310	320
gi-50913505.	SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI					
	: : ::::	:	:		: :	:
gi-21909636.	TYKVNEMEANK-DGYKTTASLKEGDSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV					
	260	270	280	290	300	310
	330	340	350	360	370	380
gi-50913505.	LFGEYTVLPLVMNKGHNNTINIYIRSTRLGLKPIGAAPALIQPRSFRLTPRSTRMKRS					
gi-21909636.	VGTLPFAVLISIVAIGGVIIYITKRKKA					
	320	330	340			

gi-50913505.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Initl: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2
>>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
initn: 31 initl: 31 opt: 62 Z-score: 60.7 expect(): 7.2
Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
(181-305:181-303)

	160	170	180	190	200	210
gi-50913505.	DITTTLTFFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR					
				:: :	::: :	:
gi-28810259.	VDVYVGNKEGGGFEPKFIVSKEQGTVDKPKPVNFNNSFATTSLKVKKNVSGN-----TGE					
	160	170	180	190	200	
	220	230	240	250	260	
gi-50913505.	LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL					
	:	: : :	: ::	: :: :	: : :	: :
gi-28810259.	LQKEFDFTLTNLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNGESIQLDKLPVGI					
	210	220	230	240	250	260
	270	280	290	300	310	320
gi-50913505.	SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI					
	: : ::::	:	:		: :	:
gi-28810259.	TYKVNEMEANK-DGYKTTASLKEGDSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV					
	270	280	290	300	310	320

FIGURE 57I

PCT/US2005/027239

330 340 350 360 370 380

gi-50913505. LFGEYTVLEPLVMNKGHNNTINIIYIRSTRLPLGLKPIGAAPALIQPRSFRLTPRSTRMKRS

gi-28810259. VGTLPFAVLISIVAIGGVIIYITKRKKA
330 340gi-50913505.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.9 expect(): 8.9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:90-143)

720 730 740 750 760

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL

gi-19224139. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
60 70 80 90 100 110

770 780 790 800 810 820

gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFSKDYKVDDEYTYTSLFNVKASDEAY

gi-19224139. VAKVISRRAGDEEKSATFKPKRLVKPIPPROPNIPKTPLPLAGEVKSLLGILSIVLLGL
120 130 140 150 160 170gi-50913505.pep
/home/morama/gas/pili/align/gi-19745305.pepgi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9
>>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.8 expect(): 9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:96-149)

720 730 740 750 760

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL

gi-19745305. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
70 80 90 100 110 120

770 780 790 800 810 820

gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFSKDYKVDDEYTYTSLFNVKASDEAY

gi-19745305. VAKVISRRAGDEEKSATFKPKRLVKPIPPROPDIPKTPLPLAGEVKSLLGILSIVLLGL
130 140 150 160 170 180gi-50913505.pep
/home/morama/gas/pili/align/orf82.pepTRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.

FIGURE 57J

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6
 >>/home/morama/gas/pili/align/orf82.pep (224 aa)
 initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
 (742-796:125-178)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDETEIL				
orf82.pep	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	100	110	120	130	140 150

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTTSPKQKGVTLTFKSDYKVDDEYTYTSLFNVKASDEAY					
orf82.pep	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLEGILSIVLLGL					
	160	170	180	190	200	210

gi-50913505.pep
 /home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9
 >>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
 initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9
 Smith-Waterman score: 52; 31.6% identity in 57 aa overlap
 (742-796:84-137)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDETEIL				
gi-21909638.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTTSPKQKGVTLTFKSDYKVDDEYTYTSLFNVKASDEAY					
gi-21909638.	VAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPLPLAGEVKSLLEGILSIVLLGL					
	120	130	140	150	160	170

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:46:18 2004

! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:02.2

! Total CPU time: 0:00:02.3

! Output File: gi-50913505.fasta

FIGURE 57K

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:====*====
50	24	5:====*=====
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:=*
60	3	2:=*=
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*
70	2	1:*=
72	0	0:
74	3	0:====
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

187/487

```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=

```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

init1 initn opt z-sc E(55) ..

```

/home/morama/gas/pili/align/gi-50913506.pep  Begin: 1  End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454 3454 3454 1016.2 4.7e-53
/home/morama/gas/pili/align/orf84.pep  Begin: 316  End: 567
! TRANSLATE of: orf84.seq check: 7868... 57 83 135 75.1 1.2
/home/morama/gas/pili/align/gi-19745307.pep  Begin: 316  End: 567
! gi|19745307|ref|NP_606443.1| protei... 57 83 135 75.1 1.2
/home/morama/gas/pili/align/gi-21909640.pep  Begin: 202  End: 524
! gi|21909640|ref|NP_663908.1| protei... 56 81 134 75.0 1.2
/home/morama/gas/pili/align/gi-28810263.pep  Begin: 316  End: 638
! gi|28810263|dbj|BAC63201.1| protein... 56 82 134 74.7 1.3
/home/morama/gas/pili/align/orf80.pep  Begin: 49  End: 352
! TRANSLATE of: orf80.seq check: 9824... 45 69 113 70.8 2.1
/home/morama/gas/pili/align/gi-19224137.pep  Begin: 25  End: 342
! gi|19224137|gb|AAL86408.1|AF447492_... 45 69 109 69.8 2.4
/home/morama/gas/pili/align/gi-19224141.pep  Begin: 277  End: 645
! gi|19224141|gb|AAL86412.1|AF447492_... 73 73 118 68.9 2.7
/home/morama/gas/pili/align/gi-21909636.pep  Begin: 44  End: 344
! gi|21909636|ref|NP_663904.1| conser... 45 98 96 66.1 3.8
/home/morama/gas/pili/align/gi-28810259.pep  Begin: 49  End: 349
! gi|28810259|dbj|BAC63197.1| hypothe... 45 98 96 66.0 3.8
/home/morama/gas/pili/align/gas15.pep  Begin: 222  End: 470
! GAS15 GAS15 42 68 96 63.8 5
/home/morama/gas/pili/align/gi-13621428.pep  Begin: 17  End: 340
! gi|13621428|gb|AAK33238.1| hypothet... 41 41 87 63.6 5.2
/home/morama/gas/pili/align/gi-19224135.pep  Begin: 193  End: 462
! gi|19224135|gb|AAL86406.1|AF447492_... 41 41 86 61.0 7
/home/morama/gas/pili/align/gi-50913505.pep  Begin: 503  End: 966
! gi|50913505|ref|YP_059477.1| Collag... 70 149 86 60.1 7.8
/home/morama/gas/pili/align/gi-13621430.pep  Begin: 60  End: 143
! gi|13621430|gb|AAK33240.1| hypothet... 43 67 67 59.2 8.7
/home/morama/gas/pili/align/gi-19745303.pep  Begin: 44  End: 344
! gi|19745303|ref|NP_606439.1| hypoth... 51 106 69 58.4 9.5
\\End of List

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

SCORES Init1: 3454 Initn: 3454 Opt: 3454 z-score: 1016.2 E(): 4.7e-53

>>/home/morama/gas/pili/align/gi-50913506.pep (556 aa)

initn: 3454 init1: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53

Smith-Waterman score: 3454; 100.0% identity in 556 aa overlap

FIGURE 58A

PCT/US05/27239
(1-556:1-556)

	10	20	30	40	50	60
gi-50913506.	MTNRRET	VREKIL	ITAKKL	MLACL	AILAVV	GLGMTR
gi-50913506.	MTNRRET	VREKIL	ITAKKL	MLACL	AILAVV	GLGMTR
	10	20	30	40	50	60
	70	80	90	100	110	120
gi-50913506.	LYKIGEG	VYNTNG	DSFINF	KYAEGV	SLTETG	PTSQEIT
gi-50913506.	LYKIGEG	VYNTNG	DSFINF	KYAEGV	SLTETG	PTSQEIT
	70	80	90	100	110	120
	130	140	150	160	170	180
gi-50913506.	NGTATYN	NARGAS	VYIAL	LLTGAT	DGR	TYPILL
gi-50913506.	NGTATYN	NARGAS	VYIAL	LLTGAT	DGR	TYPILL
	130	140	150	160	170	180
	190	200	210	220	230	240
gi-50913506.	AKSSLPS	ITKKVT	GTIDDV	NKKT	TSLSG	SVLSYS
gi-50913506.	AKSSLPS	ITKKVT	GTIDDV	NKKT	TSLSG	SVLSYS
	190	200	210	220	230	240
	250	260	270	280	290	300
gi-50913506.	TFNFNS	LTVEWK	GKMAN	ITEDG	SVMVEN	TKIGIA
gi-50913506.	TFNFNS	LTVEWK	GKMAN	ITEDG	SVMVEN	TKIGIA
	250	260	270	280	290	300
	310	320	330	340	350	360
gi-50913506.	AVVNNKA	IVGEEG	NPKNKA	EFFYS	NNPTK	GN
gi-50913506.	AVVNNKA	IVGEEG	NPKNKA	EFFYS	NNPTK	GN
	310	320	330	340	350	360
	370	380	390	400	410	420
gi-50913506.	FRKVDS	VSKT	PLIGA	IFGVYD	TSNKL	IDIV
gi-50913506.	FRKVDS	VSKT	PLIGA	IFGVYD	TSNKL	IDIV
	370	380	390	400	410	420
	430	440	450	460	470	480
gi-50913506.	LNTETYE	ITANWV	TATV	KTSAN	SKSTTY	TS
gi-50913506.	LNTETYE	ITANWV	TATV	KTSAN	SKSTTY	TS
	430	440	450	460	470	480
	490	500	510	520	530	540
gi-50913506.	DVKEAY	IESTKA	LT	DGTT	FSKSNE	GSGTV
gi-50913506.	DVKEAY	IESTKA	LT	DGTT	FSKSNE	GSGTV
	490	500	510	520	530	540
	550					
gi-50913506.	AMIGA	IGIYIV	KRRKA			
gi-50913506.	AMIGA	IGIYIV	KRRKA			
	550					

FIGURE 58B

189/487
 PCT/US2005/027239
 gi-50913506.pep
 /home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
 generated symbols 1 to: 696.
 GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
 >>/home/morama/gas/pili/align/orf84.pep (696 aa)
 initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
 Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
 (232-462:316-567)

	210	220	230	240	250
gi-50913506.	KTTSLGSVLSYSLTFELPSYTKAEVNKTVYVSDNMSEGLTFNFNSLTVEWKGMAN----				
	::: : ::: :: : : : ::: :				
orf84.pep	EKEVAVDDAELKQINSEGOQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY				
	290	300	310	320	330 340
	260	270	280	290	300 310
gi-50913506.	ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE				
	::: :: : ::: ::: : ::: ::: :: : :				
orf84.pep	IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI--VSGENQMKQIEGED				
	350	360	370	380	390
	320	330	340	350	
gi-50913506.	GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ				
	:: :: :: ::: :: : :: ::: :: : :				
orf84.pep	SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSSSEQGSGDMTIE				
	400	410	420	430	440 450
	360	370	380	390	400 410
gi-50913506.	-----IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNKYAISTQVSSGKYKIKE				
	:: : : :: : : : : : :				
orf84.pep	EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE				
	460	470	480	490	500 510
	420	430	440	450	460 470
gi-50913506.	LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF				
	: ::: : :: ::: ::: : :				
orf84.pep	TAAPDGYEIAITAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGQVIDIEEKL				
	520	530	540	550	560 570
	480	490	500	510	520 530
gi-50913506.	YSIDSRPTGNDVKEAYIESTKALTDGTFESKSNEGSGTVLLETDPNTKLGLPSTGSG				
	DEQGHSGSTTEIEDSKSSDVIIGGOGQIVETTEDTGTGMHGDGCKTEVEDTKLVQSFHF				
orf84.pep					
	580	590	600	610	620 630

gi-50913506.pep
 /home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
 >>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)
 initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
 Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
 (232-462:316-567)

FIGURE 58C

190/487

```
                210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNKTVYVSDNMSEGLTFNFNLSLTVEWKGMAN----
                |::: |::: :: |::: |:::
gi-19745307. EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                290      300      310      320      330      340

                260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAIVGEE
                |::: |::: |::: |::: |::: |:::
gi-19745307. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
                350      360      370      380      390

                320      330      340      350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
                ::| |::: |::: |::: |::: |::: |:::
gi-19745307. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQQSGDMTIE
                400      410      420      430      440      450

                360      370      380      390      400      410
gi-50913506. -----IAFRKVD SVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                ||| |:: |::: |::: |::: |::: |:::
gi-19745307. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
                460      470      480      490      500      510

                420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
                ||| |:: |::: |::: |::: |::: |:::
gi-19745307. TAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKSGSQVIDIEEKLP
                520      530      540      550      560      570

                480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTGSIG
gi-19745307. DEQGHSGSTTEIEDSKSSDVIIGGQGVIVETTEDTQTGMHGD SGCKTEVEDTKLVQSFHF
                580      590      600      610      620      630
```

gi-50913506.pep

/home/morana/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

```
SCORES  Init1: 56  Initn: 81  Opt: 134  z-score: 75.0  E(): 1.2
>>/home/morana/gas/pili/align/gi-21909640.pep (656 aa)
initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2
Smith-Waterman score: 156; 23.9% identity in 347 aa overlap
(232-547:202-524)
```

```
                210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNKTVYVSDNMSEGLTFNFNLSLTVEWKGMAN----
                |::: |::: :: |::: |:::
gi-21909640. EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                180      190      200      210      220

                260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAIVGEE
                |::: |::: |::: |::: |::: |:::
gi-21909640. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
                230      240      250      260      270      280
```

FIGURE 58D

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
          ::|          || ::|          ||::| ::| ::| ::| ::| ::| ::|
gi-21909640. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEQQQSGDMTIE
          290          300          310          320          330          340

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
          ||| || ::| || ::| ::| ::| ::| ::| ::| ::| ::| ::|
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
          350          360          370          380          390          400

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
          || || : | | : : : ||:: ||: : : :: : | | : | ||
gi-21909640. TAAPDGYEVATAITFTVNEQQQVTVNGKATKGDHIVMV-DAYKPTKGSQV-----
          410          420          430          440          450

          480          490          500          510          520          530
gi-50913506. YSIDS RPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTGSIG
          || : | :::: || : | : : : | : | : | : : | : ||
gi-21909640. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
          460          470          480          490          500

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          | : : : | ::|| : |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGOIVETT
          510          520          530          540          550          560
```

gi-50913506.pep

/home/morana/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3

>>/home/morana/gas/pili/align/gi-28810263.pep (733 aa)

initn: 82 init1: 56 opt: 134 Z-score: 74.7 expect(): 1.3

Smith-Waterman score: 155; 23.6% identity in 347 aa overlap
(232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KTTSLGSVLSYSLTFFELPSYTKKAVNKTYVSDNMSEGLTFNENSLTVEWKGMAN-----
          ||: :| :::: : : | : :| :::
gi-28810263. EKEVAVDDAELKQINSEGGQEIISVTWENQVDE--KGMAYTYSVKEVDKNGELLEPKDY
          290          300          310          320          330          340

          260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNKKAIVGEE
          | :: : : | || : | :::: : : : | : :| : : : | : || :
gi-28810263. IKKEDGLTVTNTYV--KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED-
          350          360          370          380          390

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
          ::|          || ::|          ||::| ::| ::| ::| ::| ::| ::|
gi-28810263. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEQQQSGDMTIE
          400          410          420          430          440          450

          360          370          380          390          400          410
```

FIGURE 58E

PCT/US2005/027239

gi-50913506. -----IAFRKVDVSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNKYAISTQVSSGKYKIKE
| | | | : : | | : : | : | | : : | : : : | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
460 470 480 490 500 510

gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSA NSKSTTYTSDKNKATDNSEQVGWLKNGIF
| | | : | | : : : | | : | : : : : : | | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKGSQV-----
520 530 540 550 560

gi-50913506. YSIDS RPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGE L PSTGSIG
| | : | : : : | | : | : : : | : | : | : : | : |
gi-28810263. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
570 580 590 600 610 620

gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
| : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGVIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
630 640 650 660 670 680

gi-50913506.pep
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056
generated symbols 1 to: 352.
GETSEQ from morama, September 13, 2004 17:11.

SCORES Initl: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1
>>/home/morama/gas/pili/align/orf80.pep (352 aa)
initn: 69 initl: 45 opt: 113 Z-score: 70.8 expect(): 2.1
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap
(284-556:49-352)

gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVN-NKAIVGEE
: : | | : | : | : : : : : : | : |
orf80.pep. ATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEV LMPETAFTFTIEPDMTASGKE
20 30 40 50 60 70

gi-50913506. GNPN-KAEFFYSNNPTKGN TYDNLDKPKDNGITTSKEDSKIVYTYQIAFRKVDVSVSKTP
| : : | : : : : | | | : | : | : : : | : : : : | : :
orf80.pep. GSLDIKNGIVEGLDKQVTVKYKNTDKTSQTK-IAQFDFSKVKFPATGVYRYMVSEKNDK
80 90 100 110 120 130

gi-50913506. LIGAIFGVYDTSNKLIDIVTTNKN-----YAISTQ-VSSGKYKIKELKAPKGYSLN
| : : | : : : : | : : : : | : : : : | : : : : | : :
orf80.pep. KDGI---TYDDKKWTVVDVYVG NKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
140 150 160 170 180 190

gi-50913506. TETYEITANW-----VTATVKTSA NSKSTTYTSDKNKATDNSEQVG-----WLKNGI
| : : | : : | : : : : : : : : : : | : |
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
200 210 220 230 240 250

470 480 490 500 510

FIGURE 58F

PCT/US05/27239

gi-50913506. FYSIDSRPTGNDV-----KEAYI-----ESTKALTDGTTFSKSNESGTVLLETDI
::: ||: ||:: | : | : | : | : | :
orf80.pep SVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNQKTDESADIVV
260 270 280 290 300 310

520 530 540 550
gi-50913506. PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
| : : | | : | | : | : | : | : | : | :
orf80.pep TNKRDTQVP-TGVVGTLPFAVL SIVAIGGV-IYITKRKKA
320 330 340 350

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4
>>/home/morama/gas/pili/align/gi-19224137.pep (342 aa)
initn: 69 initl: 45 opt: 109 z-score: 69.8 expect(): 2.4
Smith-Waterman score: 169; 26.0% identity in 334 aa overlap
(257-556:25-342)

230 240 250 260 270 280
gi-50913506. NKTVYVSDNMSEGLTFNFNLSLTVIEWKKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFI
| : : : | : : : | : : | :
gi-19224137. MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITN-FN----
10 20 30 40

290 300 310 320 330 340
gi-50913506. YDSLESISPNI SYKAVVN-NKAIVGEEGN-PNKAIEFFYSNNPTKGNTYDNLDKPKDKGNG
| : | : | : | : | : | : | : | : | : | :
gi-19224137. DDTL--LMPKTDYTF SVNPDSAATGTESNLPIKPGIAVNNQDIK-VSYSNTDKTSGKEKQ
50 60 70 80 90 100

350 360 370 380 390 400
gi-50913506. ITSKEDESKIVYTYQIAFRKVDSVSKTPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQV
: : | : : : | : : | : | : | : | : | : : :
gi-19224137. VVV-DFMKVTFPSVGIYRYVVTENK---GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKY
110 120 130 140 150 160

410 420 430 440 450
gi-50913506. SSGKYKIKELKAPKGY--SLNTETYELTANWVTATVKTSANSKSTTYTSDKNKATDNS--
: | : | : | : | : | : | : | : | : | :
gi-19224137. IVSKKGD SATKEPIQFNNSFETTSLKIEKE-VTGNIGDHKKAF TTTLTLPNEYEASSV
170 180 190 200 210 220

460 470 480 490
gi-50913506. ---EQVGWLKN---GIFYSI---DSR-----PTGND--VKEAYIE-----STKALTDG
| : | : | : | : | : | : | : | : | : | :
gi-19224137. VKIEENGQTKDVKIGEAYKFTLND SQSVILSKLPVGIN YKVEEAEANQGGYTTTATLKD G
230 240 250 260 270 280

500 510 520 530 540 550
gi-50913506. TTFSKSNEG----SGTVLLETDI PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVK
: | | : : | : | : | : | : | : | : | :
gi-19224137. EKLSTYNLGQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLPFAVL SIVAIGGV-IYITK
290 300 310 320 330

gi-50913506. RRKA

| : |

FIGURE 58G

PCT/US05/27239

194/487

gi-19224137. RKKA
340

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 73 init1: 73 opt: 118 Z-score: 68.9 expect(): 2.7
Smith-Waterman score: 174; 23.9% identity in 406 aa overlap
(115-483:277-645)

90	100	110	120	130	140
gi-50913506.	VSLTETGPTSQEITTIANGINTGKIKPFSTENV	SISNGTATYNARGASVY--IALLTGAT			
gi-19224141.	IYTFDYIAGLDKVQLSAELSLFLENKEVLENTS	SISNFKSTIGGQEITYKGTVNVLYGNE			
250	260	270	280	290	300
150	160	170	180	190	
gi-50913506.	DGRTYNPILLAASYNGEGNLVTKNIDSKSNYLYG	QTSVAKSSLPSITKKVTG-----T			
gi-19224141.	STKESNYITNGLSNVG-GSIESYNTETGEFVWYV	YVNPNTNIPYATMNLWGFGRARSNT			
310	320	330	340	350	360
200	210	220	230	240	250
gi-50913506.	ID---DVNKKTTSLGSVLSYSLTF--ELPSYTK	EAVNKTIVYVSDNMSEGLTFNFNSLTVE			
gi-19224141.	SDLENDANTSSAELGEIQVYEVPEGEKLPSSYG	VDVTKLTLRTD-ITAGLNGFQ-----			
370	380	390	400	410	
260	270	280	290	300	310
gi-50913506.	WKGKMANITEDGSVMVENTKIGIAKEVNNGFN	LSFIYDSLESISPNISYKAVVNNKAIVG			
gi-19224141.	-----MTKRQRIDFG--NNIQNKAFIIKV--	TGKTDQSGKPLVVQSNLAS			
420	430	440	450	460	
320	330	340	350	360	
gi-50913506.	EEGNPNKAEFFYSNNPTKGNTY--DNLDKPKD	KNGITSKEDSKIIVYTY-----QIAF			
gi-19224141.	FRGASEYAAF-----TPVGGNVYFQNEIAL	SPSKGSGSGKSEETKPSITVANLKRVAQLRF			
470	480	490	500	510	
370	380	390	400	410	420
gi-50913506.	RKVDVSVSKTPLIGAIFGVYDTSNKLIDI-VT	TNKNGYAISTQVSSGKYKIKELKAPKGY			
gi-19224141.	KKM-STDNVPLPEAAFELRSSNGNSQKLEASS	NTQGEVHFKDLTSGTYDLYETKAPKGYQ			
520	530	540	550	560	570
430	440	450	460		
gi-50913506.	-----LNTETYEIT-----ANWVT--	ATVKTSANSKSTTYTSDKNKATDNSEQV	GWLKN		
gi-19224141.	QVTEKLATVTVDTTKPAEEMVTWGS	PHSSVKVEAN-KEVTIVNHKETLTFSGK	KI-WEND		
580	590	600	610	620	630
470	480	490	500	510	520
gi-50913506.	GIFYSIDSRPTGNDVKEAYIESTKALTDG	TTFSKSNEGSGTVLLETDPNTKLGELP	STG		
gi-19224141.	----RPDQRP	AKIQVQLLQNGQKMPNQIQEVTKDND	WSYHFKDLPKYDAKNQ	EYKYSVEE	
640	650	660	670	680	

FIGURE 58H

gi-50913506.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
 (298-556:44-344)

	270	280	290	300	310	320
gi-50913506.	NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT					
	::: : ::: : ::: :					
gi-21909636.	ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA					
	20	30	40	50	60	70
	330	340	350	360		
gi-50913506.	KGNTYDNLDKKPKDNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK					
	: : : : : : : :: : :					
gi-21909636.	SGKTKDGLLEIKPGIVNGLTEQIIISYTNTPDKPDSKVST-EFDFSKVVFPFGIGVYRYTVSE					
	80	90	100	110	120	
	370	380	390	400	410	420
gi-50913506.	TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY					
	: : :: : : : : :: : : : ::: : : : :					
gi-21909636.	KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEOGTDVKKPVPNFNNSFATTSL					
	130	140	150	160	170	180
	430	440	450	460		
gi-50913506.	EITANWVTATVKTSAN-----SKSTTYTSDK---NKATDNSE-QVGW-----LKNGI					
	:: : : : : :: : : : ::: : : :					
gi-21909636.	KVKKNVSGNTGELQKEFDFTLTNLTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNGE					
	190	200	210	220	230	240
	470	480	490	500	510	
gi-50913506.	FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----					
	:: : : : : : : :: : : :					
gi-21909636.	SIQLDKLPVGITYKVNEMEANK---DGYKTASLKEGDGQSKMYQLDMEQKTDESADIV					
	250	260	270	280	290	300
	520	530	540	550		
gi-50913506.	IPNTKLGELPSTGSIGTYLEKAIGSAAMIGAIGTYIVKRRKA					
	: : : : : : : : : : : :: :					
gi-21909636.	VTNKRDTQVP-TGVVGTLPFAVL SIVAIGGV-IYITKRKKA					
	310	320	330	340		

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
 (298-556:49-349)

FIGURE 58I

PCT/US2005/027239

```
270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: : |:: :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
20          30          40          50          60          70

330      340      350      360
gi-50913506. KGNTFYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
:|:| |:|: || |:| :| |||: |::| || :||:
gi-28810259. SGKTKDGLLEIKPGIVNGLTEQIISYTNTPDKPDSKVKST-EFDFSKVVFPGIGVYRYTVSE
80          90          100         110         120         130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
| : |: |||:: :|: : ||:| :: : :| : ::| |:: |:| :
gi-28810259. KQ--GDVEGITYDTKKWTVDVVVGNGKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
140         150         160         170         180         190

430      440      450      460
gi-50913506. EITANWVTATVKTSAN-----SKSTTYTSDK----NKATDNSE-QVGW-----LKNGI
:::| |::: :||::: |:| :|:::| :|||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKNEKFVVKIGTPYKFKLKNGE
200         210         220         230         240         250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
::|: |:| | :|::| || :| :||:| :
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIV
260         270         280         290         300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
:| : ::| ||:| | :| : ||:: |||:|:|
gi-28810259. VTNKRDTQVP-TGVVGTLPFAVL SIVAIGGV-IYITKRKKA
310         320         330         340
```

gi-50913506.pep
/home/morana/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5
>>/home/morana/gas/pili/align/gas15.pep (762 aa)
initn: 68 initl: 42 opt: 96 z-score: 63.8 expect(): 5
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap
(283-535:222-470)

```
260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLES-ISPNI SYKA---VVNN--K
                ||:: ::|:: |:|::| |::
gas15.pep     VWYYSNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQL
200         210         220         230         240         250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTFYDNLDKKPKDKNGITSKEDSKIVYTYQIAFRK--V
:| | : :| : |:| :| : | | | : :: : | : :|| :
gas15.pep     SIFESEDKGDKYNKGYQNLLSGGLVPT---KPPTPGDPPMPNPQPQ---TTSVLIRKYAI
260         270         280         290         300

370      380      390      400      410      420
```

FIGURE 58J

PCT/US05/27239

gi-50913506. DSVSKTPLIGAIFGVY-DTSNKL-IDIVTTNKNKYAISTQVSSGKYKIKELKAPKGYSLN
: || | || : : | : | : : : | | | : : | | : || : | || :
gas15.pep GDYSKL-LEGATLQLTGDNVNSFQARVFSSNDIGERI--ELSDGTYTLTELNSPAGYSIA
310 320 330 340 350 360

gi-50913506. TE-TYEITANWVTATV--KTSANSKSTTYTSDKNKATDNSEQVGWLKN--GIFYSIDSR
| : : | : | : : | : : : : | : : | : : | : : | : :
gas15.pep EPITFKVEAGKVYTIIDGKQIENPNKEIVEPYSVEAYNDFEFSVLTTQNYAKFYAKNK
370 380 390 400 410 420

gi-50913506. PTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDPNTKLGEPLSTGSIGTYLFKA
: : : | : : | : | : : | : | : | : | : | : | : | :
gas15.pep NGSSQVVYCFNADLKSPPD-----SEDGCKTMT-----PDFTTGEVKYTHIAGRDLFKY
430 440 450 460 470

gi-50913506. IGSAAMIGAIGIYIVKRRKA
540 550

gas15.pep TVKPRDTPDPTFLKHIKKVIEKGYREKGOAIEYSGLTETQLRAATQLAIYYFTDSAELDK
480 490 500 510 520 530

gi-50913506.pep

/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 41 Initn: 41 Opt: 87 z-score: 63.6 E(): 5.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 41 init1: 41 opt: 87 Z-score: 63.6 expect(): 5.2
Smith-Waterman score: 109; 22.6% identity in 345 aa overlap
(256-556:17-340)

gi-50913506. VNKTVYVSDNMSEGLTFNFNLSLTFVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSF
| | | :
gi-13621428. MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTK-----NLIDL
10 20 30 40

gi-50913506. IYDSLESISPNTISYKAVVNNKATVGEENPNKAEFFYSNNPTKGN TYDNLDKKPKDKNGI
:
gi-13621428. VNSN--ALIPNTDTEFKIEPDITVNEEDGNKFKGVAL--NTPMTKVITYNSDK--GGSNTK
50 60 70 80 90

gi-50913506. TSKED-SKI-----VYTYQIAFRKVDV-----SKTPLIGAIF
| : : | : : | :
gi-13621428. TAEFDFSEVTFEKPVGYYKYVTEEKIDKVPVSYDTTSYTVQVHVLWNEEQQKPVATYIV
100 110 120 130 140 150

gi-50913506. GGYDTSNKLIDIVTT-NKNGYAISTQVS-SGKYKIKELKAPKGYSLNTETYEITANWVTA
| : | :
gi-13621428. GYKEGSKVPIQFKNSLDSTTLTVKKVSGTGGDRSKDFNF--GLTLKANQYYKASEKVM
160 170 180 190 200 210

gi-50913506. TVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGND--VKEAYIESTKAL
| :
440 450 460 470 480 490

FIGURE 58K

500	510	520	530	540	550
gi-50913506. SKSNEGSGTVLLETDPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA					
: :: : : : : : : :					
gi-19224135. PESEDGGGTI--DPDISTMK--EVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKG					
440	450	460	470	480	

FIGURE 58L

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTD LKTLKTYNNGKGYHGFESMDERTLAVTK
490 500 510 520 530 540

gi-50913506.pep

/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8
>>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
initn: 149 init1: 70 opt: 86 Z-score: 60.1 expect(): 7.8
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(33-428:503-966)

gi-50913506.	NRRET	VREKIL	ITAKKL	MLACL	AILAV	VGLGM	TRVS-	ALSKDD	TAQLK	ITNIE	GGPT	VTTL
gi-50913505.	YSQWV	YHGTF	DYSSY	QGGSF	NRGQI	HYRYR	GIVSV	SDGIR	RDDAV	KNSLL	GVNG-	L
	480	490	500	510	520							
gi-50913506.	YKIGEG	VYNTNG	DSFIN	FK----	YAEGV	SLTET	GPTSQ	EIT-	TIANG	INTGK	IKPF	STEN
gi-50913505.	LQRFV	NINPEN	KLSVIG	FQGSAD	YHAGK	WYPDQ	SPRGG	FYQPN	LNNSR	DAELL	KGWST	NS
	530	540	550	560	570	580						
gi-50913506.	VSISNG	TATYN	ARGAS	VYIAL	LTGAT	-----	DGRTY	NPILL	AAS----	YNGEG	NLVTK	
gi-50913505.	LLDPNT	LTALH	NNGTNY	HAALL	KAKEIL	NEVKDD	GRRKI	MFISD	GVPTF	YFGED	GYRSG	
	590	600	610	620	630	640						
gi-50913506.	NIDS--	KSNLY	YGQ--	TSVA---	KSSLP	SITKK	VGTID	DVNKK	TSLGS	VLSS	SLTFE	
gi-50913505.	NGSSN	DRNNV	TRSQ	EGSKL	AIDFKA	RYPNL	SIYSL	GVSKD	INS	SDTAS	SSPVV	LKYL
	650	660	670	680	690	700						
gi-50913506.	LPSY--	-----	TKEAV	NKTVY	-----	VSDNM	SEGLT	F--	NFNSL	TVEWK	GRMAN	ITE
gi-50913505.	EHYYG	ITDTA	ELEK	TENK	IVEDS	KLSQL	GISDS	LSQYV	VDYDK	OPDVL	VTRK	SKVND
	710	720	730	740	750	760						
gi-50913506.	-----	-----	DGSVM	VENTK	IGIAKE	VNNGF	NLSFI	YDS--	LESIS	PNI---		
gi-50913505.	ILYQK	DQVQE	AGKDI	IDKVV	FTPKT	TTSQPK	GKVTL	TFKSD	YKVD	DEYTY	TLSFN	VKAS
	770	780	790	800	810	820						
gi-50913506.	SYKAV	VNNKA	IVGEE	GNPNK	AEFFY	SNNPT	KGNTY	DNLDK	KPKD	-GNGI	TSKED	SK----
gi-50913505.	AYEKY	KDNEG	RYSEM	GSDTD	YGTN	QTS	SGKGL	PSNSD	ASVNY	MADGR	EQKL	PYKHP
	830	840	850	860	870	880						
gi-50913506.	IVYTY	QIAFR	KVDS-	VSKTP	LIGAIF	GVYDT	SNKLI	-DIVT	TNKN	GYAIS	TQVSS	GK-YK

FIGURE 58M

FIGURE 58N

gi-19745303. K G K T K B G L D I K P G V I D G L E N T K T I H Y G N S D K T T A K E K S V N F D F A N V K F P G V G V Y R Y T V S E
 80 90 100 110 120 130

gi-50913506. L I G A I F G V - Y D T S N K L I D I V T T N K N - - G Y A I S T Q V S S - G K Y K I K E L K A P K G Y S L N T E T Y E
 : | : | : : : | : : : | : : | : : | : : | : : : | : :
 gi-19745303. V N G N K A G I A Y D S Q Q W T V D V Y V V N R E D G G F E A K Y I V S T E G G Q S D K K P V L F K N F - F D T T S L K
 140 150 160 170 180

gi-50913506. I T A N W V T A T V K T S A N S K S T T Y T S D K N K A T D N S E Q V G W L K N G I F - - - - - Y S I D S R - - -
 : | : | : : : | : | : | : : : | : : : : | : | : : | : : :
 gi-19745303. V T K K - V T G N T G E H Q R S F S F T L L L T P N E C F E K G Q V V N I L Q G G E T K K V V I G E E Y S F T L K D K E
 190 200 210 220 230 240

gi-50913506. - - - - - P T G - - - - - N D V K E A Y I E S T K A L T D G T T F S K S N E G - S G T V L L E T D - - - I P N T
 : | : | : : : : : | : : : | : | : : | : | : : | : : :
 gi-19745303. S V T L S Q L P V G I E Y K V T E E D V T K D G Y K T S A T L K D G D V T D G Y N L G D S K T T D K S T D E I V V T N K
 250 260 270 280 290 300

gi-50913506. K L G E L P S T G S I G T Y L F K A I G S A A M I G A I G I Y I V K R R K A
 : : : | | : | | : | : | : : | : : | : : | : : | : : | : : | : :
 gi-19745303. R D T Q V P - T G V V G T L A P F A V L S I V A I G G V - I Y I T K R K K A
 310 320 330 340

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:45:54 2004

! Completion time: Wed Sep 15 18:46:02 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:01.9

! Total CPU time: 0:00:02.0

! Output File: gi-50913506.fasta

FIGURE 580

!!SEQUENCE_LIST 1.0 05/27/239

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

T0: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	1	5:= *
46	7	5:====*==
48	21	5:====*=====
50	7	5:====*==
52	2	4:== *
54	4	3:==*=
56	3	3:==*
58	0	2: *
60	4	2:==*
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:===
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59


```
102 PCT/US05/27239
104 0 0:
106 0 0:
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=
```

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

init1 initn opt z-sc E(55)...

```
/home/morana/gas/pili/align/gi-13621430.pep Begin: 1 End: 215
! gi|13621430|gb|AAK33240.1|hypothet... 1338 1338 1338 233.9 1.8e-09
/home/morana/gas/pili/align/gi-19745305.pep Begin: 1 End: 193
! gi|19745305|ref|NP_606441.1|hypoth... 163 243 273 82.2 0.5
/home/morana/gas/pili/align/gi-28810261.pep Begin: 2 End: 187
! gi|28810261|dbj|BAC63199.1|hypoth... 164 239 268 81.5 0.55
/home/morana/gas/pili/align/gi-19224139.pep Begin: 2 End: 187
! gi|19224139|gb|AAL86410.1|AF447492... 164 236 265 81.0 0.57
/home/morana/gas/pili/align/orf82.pep Begin: 30 End: 222
! TRANSLATE of: orf82.seq check: 4296... 163 235 264 81.0 0.58
/home/morana/gas/pili/align/gi-21909638.pep Begin: 2 End: 181
! gi|21909638|ref|NP_663906.1|hypoth... 164 239 261 80.5 0.62
/home/morana/gas/pili/align/gi-19745303.pep Begin: 84 End: 183
! gi|19745303|ref|NP_606439.1|hypoth... 121 121 126 61.4 6.7
/home/morana/gas/pili/align/gi-13621428.pep Begin: 6 End: 174
! gi|13621428|gb|AAK33238.1|hypothet... 58 86 122 60.9 7.2
/home/morana/gas/pili/align/gi-19224137.pep Begin: 93 End: 201
! gi|19224137|gb|AAL86408.1|AF447492... 88 88 119 60.4 7.5
/home/morana/gas/pili/align/gi-50913503.pep Begin: 549 End: 625
! gi|50913503|ref|YP_059475.1|Fibron... 73 73 117 60.4 7.6
/home/morana/gas/pili/align/gi-19224134.pep Begin: 631 End: 697
! gi|19224134|gb|AAL86405.1|AF447492... 73 73 115 60.1 7.8
\\End of List
```

gi-13621430.pep

/home/morana/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1|hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09

>>/home/morana/gas/pili/align/gi-13621430.pep (215 aa)

initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09

Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap
(1-215:1-215).

```
10 20 30 40 50 60
gi-13621430. MKKSILRLAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-13621430. MKKSILRLAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
10 20 30 40 50 60
70 80 90 100 110 120
gi-13621430. EALDKESPLPNSVTTSVKGNKGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
gi-13621430. EALDKESPLPNSVTTSVKGNKGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
```

FIGURE 59A

FIGURE 59B

(8-213:2-187) 1505/27239

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-28810261.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::| |::| |::|
gi-28810261.  ESIDAMKTIEE---ITIAGSGKASF SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSSELIFKQYSEKTPEPHQPDTEKEKPQKKRNGI
      :|| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-28810261.  VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::: ::| : |||:
gi-28810261.  LPLAGEVKSLGLGILSIVLLGLLVLLYV-KKLKSRL
      160     170     180
```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```

SCORES   Init1: 164   Initn: 236   Opt: 265   z-score: 81.0   E(): 0.57
>>/home/morana/gas/pili/align/gi-19224139.pep   (189 aa)
   initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303;   30.6% identity in 206 aa overlap
(8-213:2-187)
```

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-19224139.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::| |::|
gi-19224139.  ESIDAMKTIEE---ITIAGSGKASF SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSSELIFKQYSEKTPEPHQPDTEKEKPQKKRNGI
      :|| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-19224139.  VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::: ::| : |||:
gi-19224139.  LPLAGEVKSLGLGILSIVLLGLLVLLYV-KKLKSKL
      160     170     180
```

FIGURE 59C

FIGURE 59D

PCT/US05/27239

20

30

40

gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEV
 |::| : : : : : |::|::| |::|::| |::|::| : : |::|::| : : |::|::|
 gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
 50 60 70 80 90

gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQYSEKTPPEPHQPDTEKEKPQKRNIGI
 :|| |::| :| ::::| :| :|| :|| :| :|::| :| :|
 gi-21909638. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPROPNI-----PKTP-----
 100 110 120 130 140

gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
 || :||: | :| :||::: :| :|::| :|
 gi-21909638. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
 150 160 170 180

gi-13621430.pep

/home/morana/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7
 >>/home/morana/gas/pili/align/gi-19745303.pep (344 aa)
 initn: 121 init1: 121 opt: 126 z-score: 61.4 expect(): 6.7
 Smith-Waterman score: 126; 27.0% identity in 100 aa overlap
 (59-155:84-183)

gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTSVKGNG-KTSF
 |::|::| : : : | | |::| : : : |
 gi-19745303. LMPKADYTFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDF
 60 70 80 90 100 110
 gi-13621430. EQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNNKLGETE
 : : | || |::| : : : |::: || : | :||: |::| :||::| ::::
 gi-19745303. ANVKFPGVGVRYTVSEVNGNKAGIAYDSQQWTVDVYVNRDGGFEAKYIVSTEGGQSD
 120 130 140 150 160 170
 gi-13621430. KSELIFKQYSEKTPPEPHQPDTEKEKPQKRNIGILPSTGEMVSYVSALGIVLVATITLY
 |: :::|::| : : :
 gi-19745303. KKPVLKFNFFDTSLSKVTKKVTGNTGEHQRSFSFTLLLPNECFEKGQVVNILOGGETKK
 180 190 200 210 220 230

gi-13621430.pep

/home/morana/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2
 >>/home/morana/gas/pili/align/gi-13621428.pep (340 aa)
 initn: 86 init1: 58 opt: 122 z-score: 60.9 expect(): 7.2
 Smith-Waterman score: 135; 29.1% identity in 172 aa overlap

FIGURE 59E

FIGURE 59F

PCT/US05/27239

SCORES Initl: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6
 >>/home/morana/gas/pili/align/gi-50913503.pep (627 aa)
 initn: 73 initl: 73 opt: 117 Z-score: 60.4 expect(): 7.6
 Smith-Waterman score: 118; 28.7% identity in 87 aa overlap
 (129-215:549-625)

100	110	120	130	140	150
gi-13621430.	KIHQLLGKNSQYHYDET	VYEVVIYVLYNEQ	SGALETNLVS	NKLGETEKSELIF	KQEYSEK
gi-50913503.	IETEDTKEPEVLMGGQ	SESVEFTKDTQT	GMSGFSETATVV	-----	EDTRPKLVFHF
520	530	540	550	560	570
160	170	180	190	200	210
gi-13621430.	TPEPHQPDTEKEKE	PQKRN	GILPSTGEMVS	YVSALGIVLV	ATITLYSIYK
gi-50913503.	KVEEN-----	REKPTKNITPIL	PATGDIENVLA	FLGILILSVLS	SIFSLLKNKQ
580	590	600	610	620	

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492_2 protein F [Streptococcus pyogenes]

SCORES Initl: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8
 >>/home/morana/gas/pili/align/gi-19224134.pep (698 aa)
 initn: 73 initl: 73 opt: 115 Z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 115; 27.4% identity in 73 aa overlap
 (143-215:631-697)

120	130	140	150	160	170
gi-13621430.	DETVYEVVIYVLYNEQ	SGALETNLVS	NKLGETEKSELIF	KQEYSEKTP	EPHQPDTEKEK
gi-19224134.	VLMGGQSESVEFTKDT	QTGMSGFSETV	TIVEDTRPKLV	FHF	DNNEPKVEEN-----
610	620	630	640	650	
180	190	200	210		
gi-13621430.	PQKRN	GILPSTGEMVS	YVSALGIVLV	ATITLYSIYK	KLKTSK
gi-19224134.	PTKNITPILPATG	DIENVLAFLG	ILILSVLSIFS	LLKNKQNNKV	
660	670	680	690		

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:45:36 2004

! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:00.3

! Total CPU time: 0:00:00.4

! Output File: gi-13621430.fasta

FIGURE 59G

TYPE 3 pilus motifs
protein F2 like fibronectin-binding protein
Length: 696-733
LPXTG
pilin motif consensus PK (X₇) K
E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

TYPE 4 pilus motifs
protein F2 like fibronectin-binding protein
Length: 1161
LPXTG
pilin motif consensus PK (X₇₋₈) K
E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYYQOVTEK
		156 PKMSVVSKYCK
		674 PKYDAKNQEYK
	E boxes	563 YDLYETKAPKGY
		940 YTFVETAAPDGY

FIGURE 60

Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

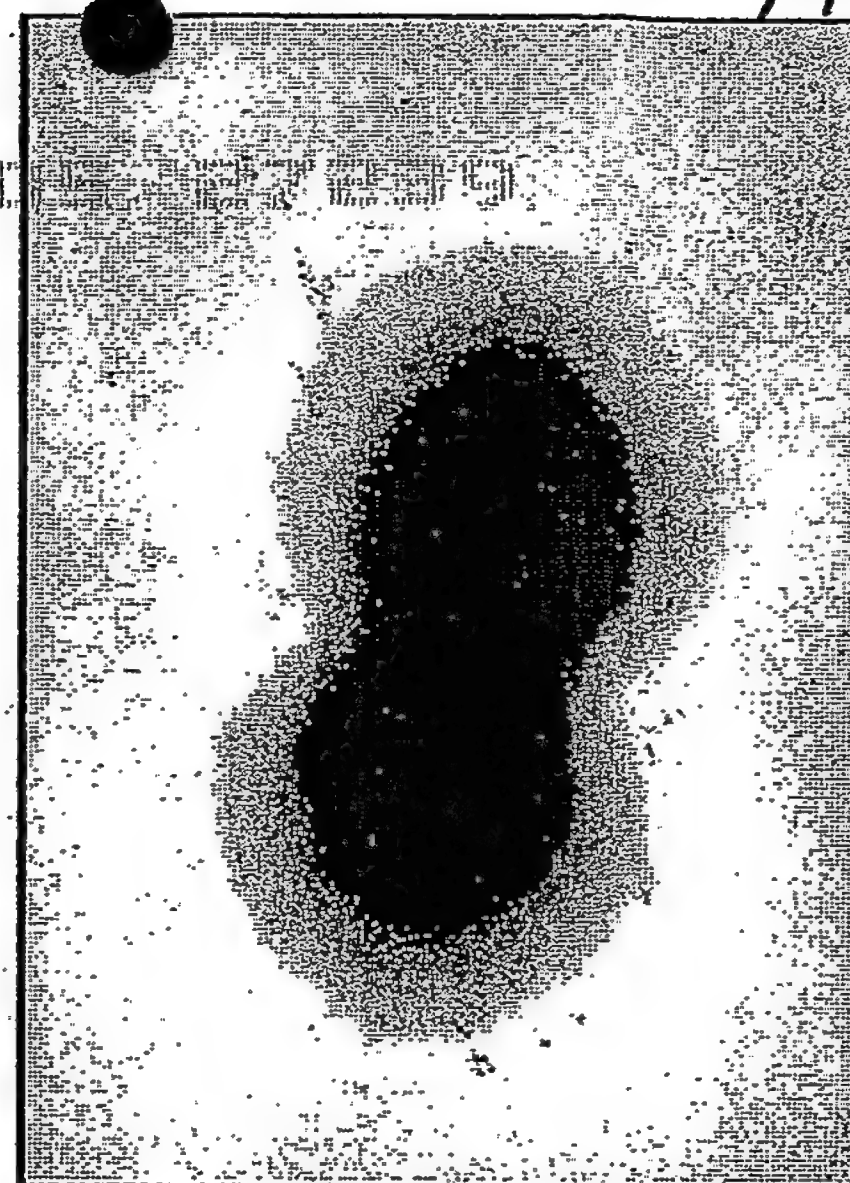
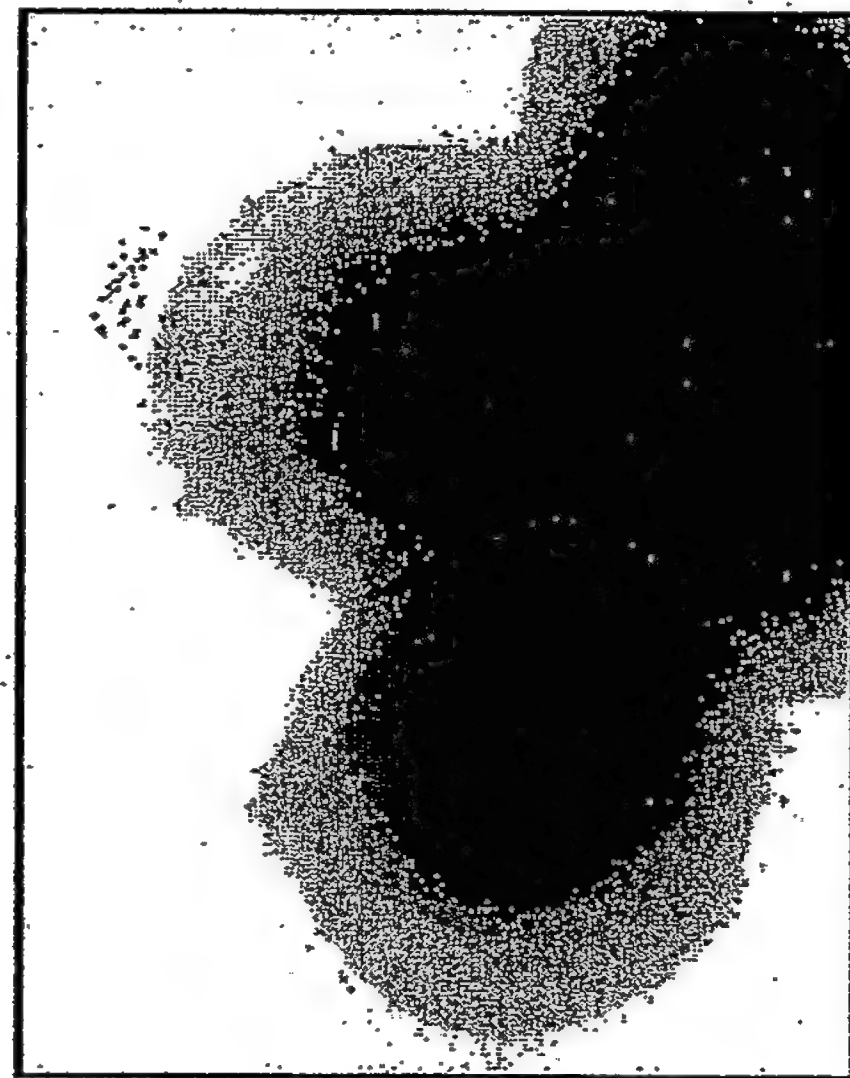
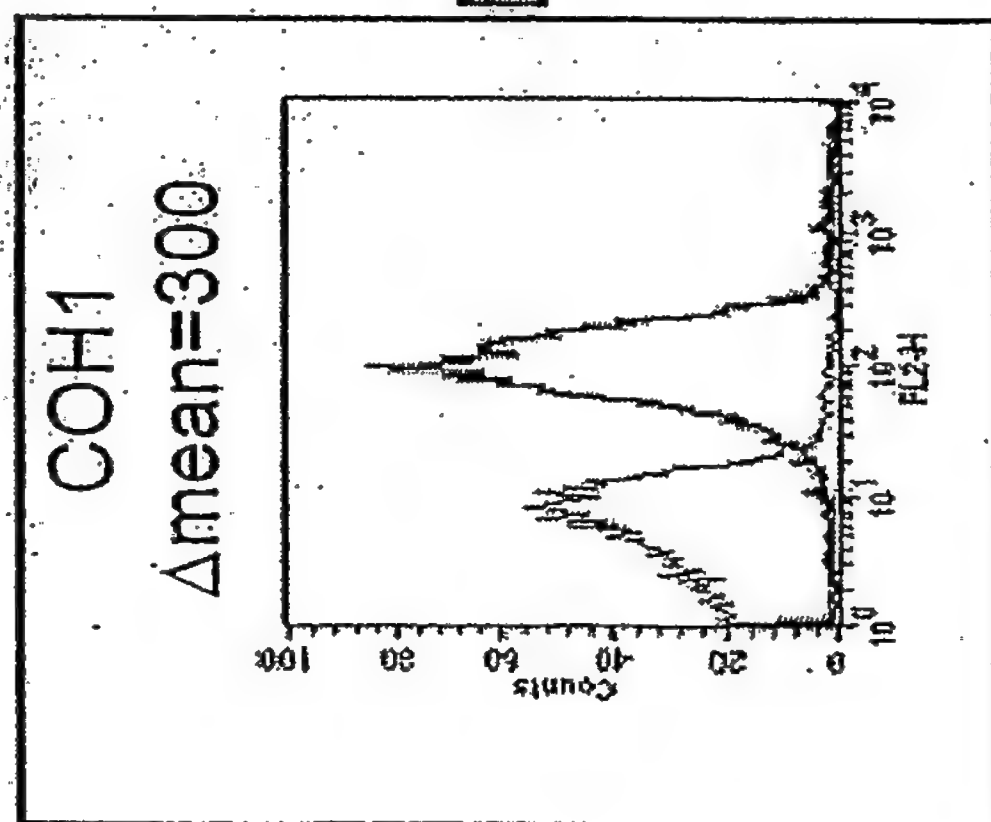
WO 2006/078318

PCT/US2005/027239

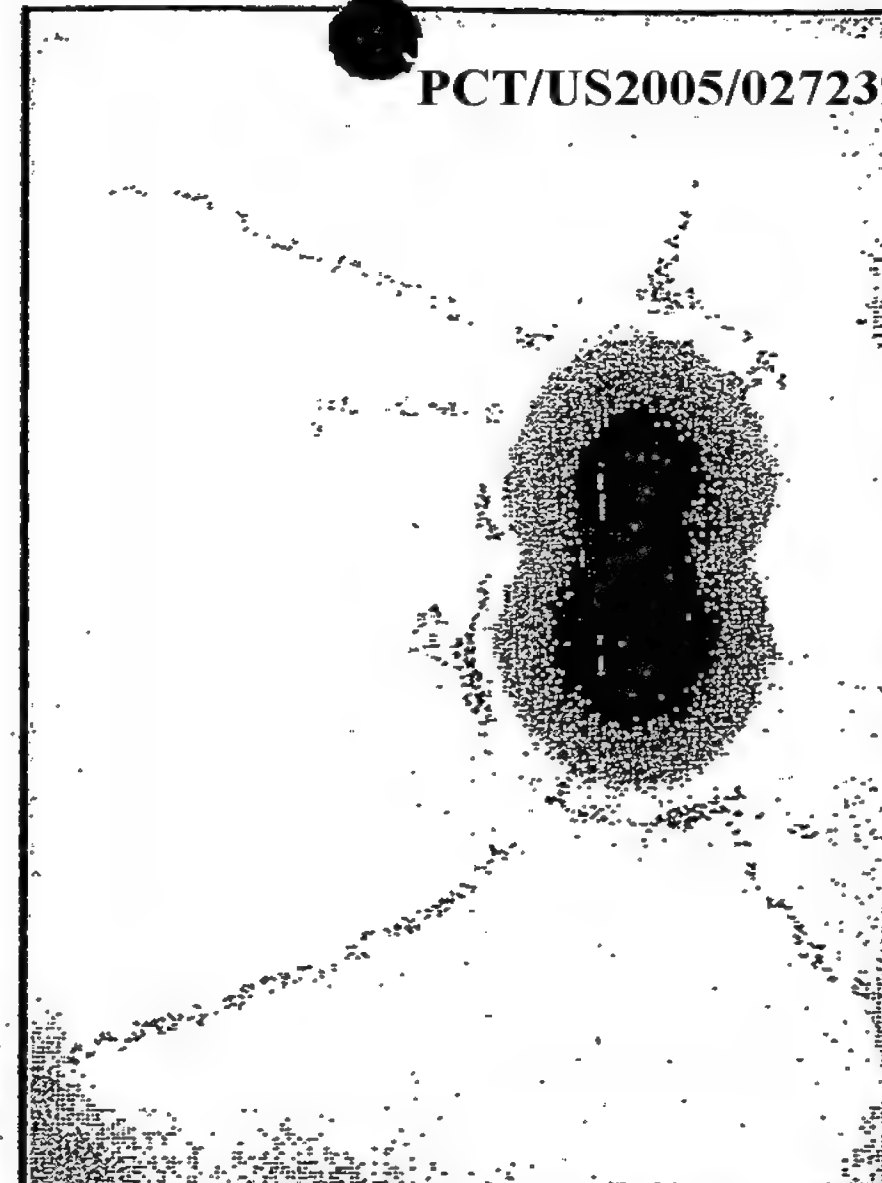
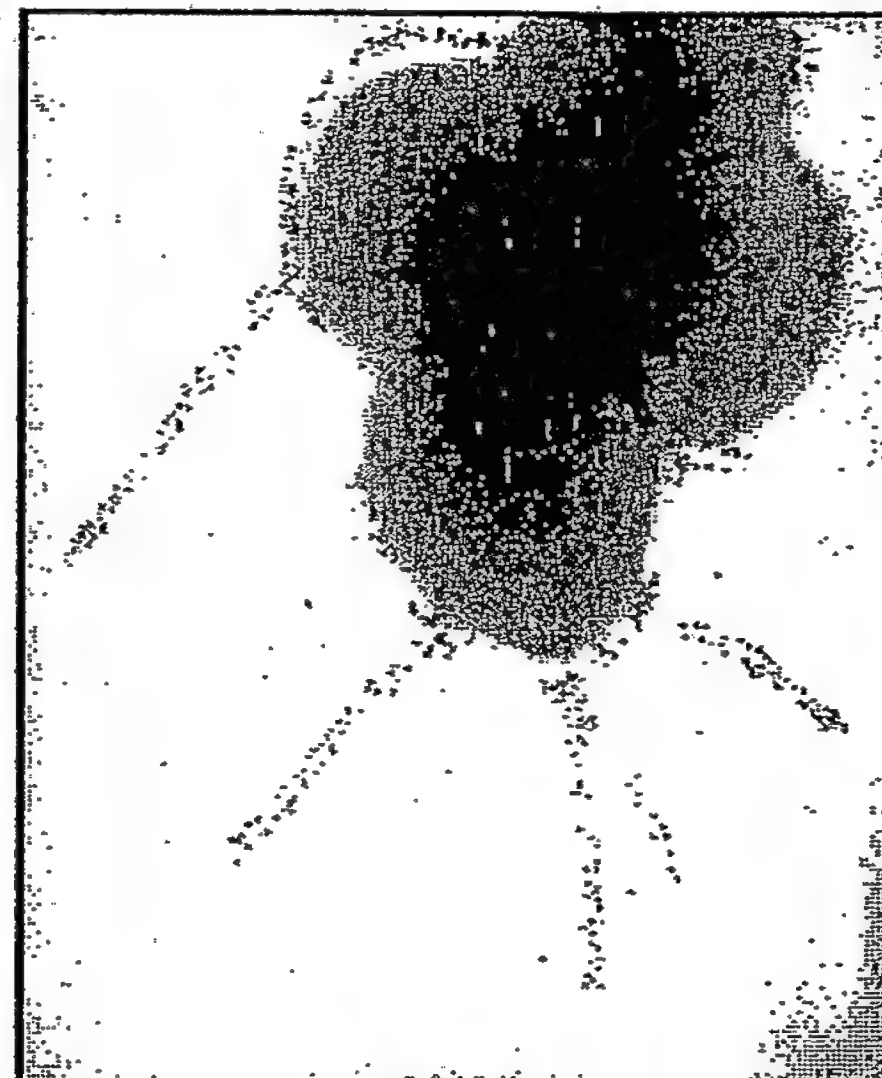
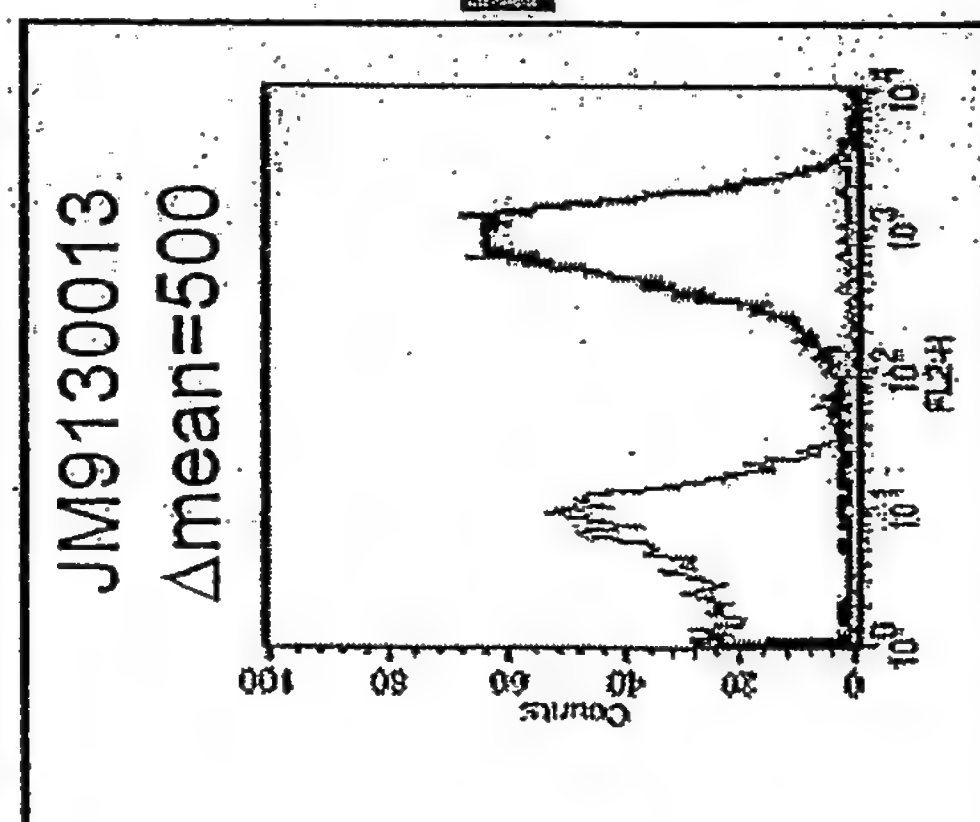
2/1/487

α -GBS80

Immunogold Electron Microscopy



α GBS80



PCT/US2005/027239

CHIRON VACCINES

FIGURE 61

Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

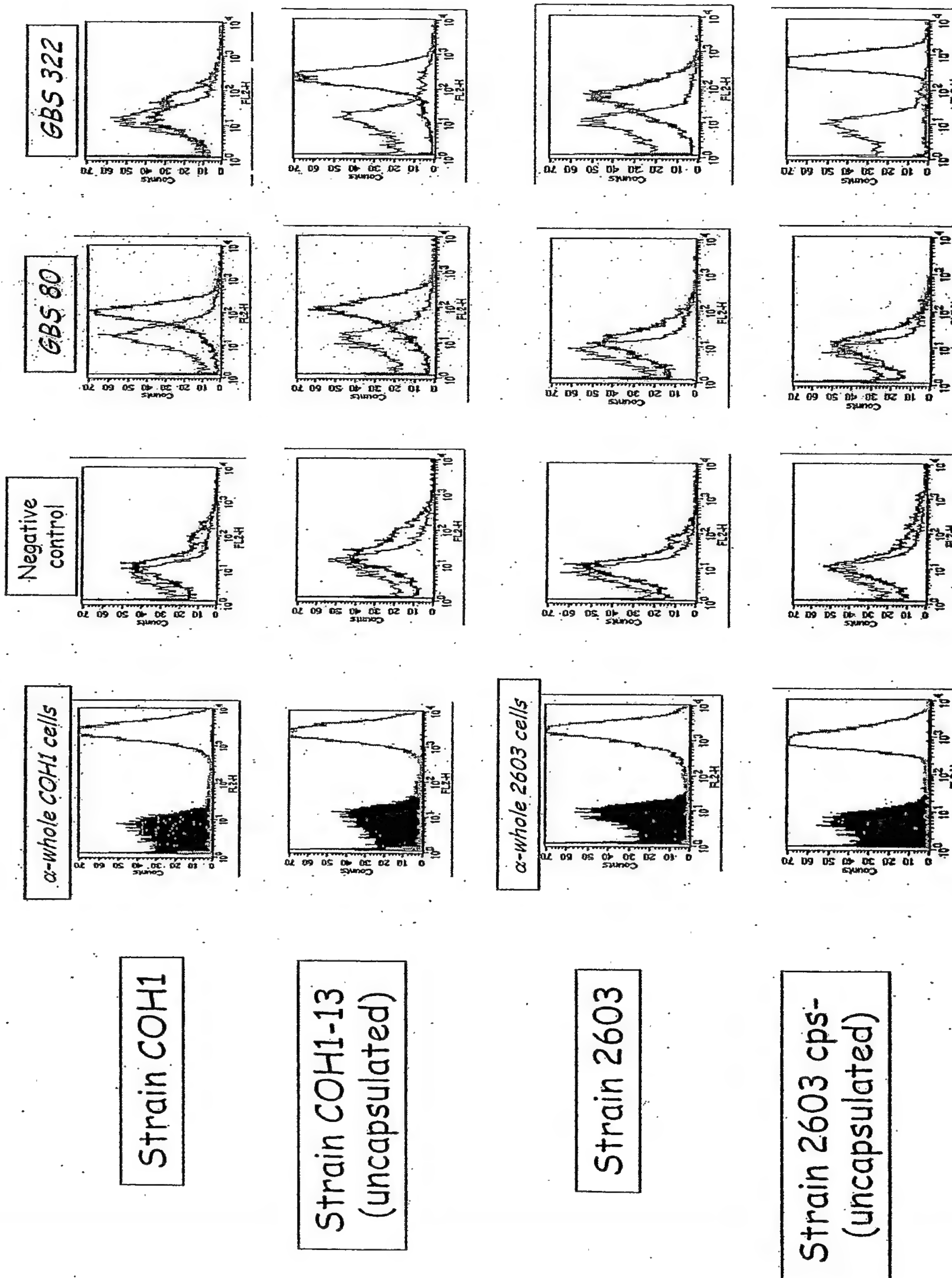
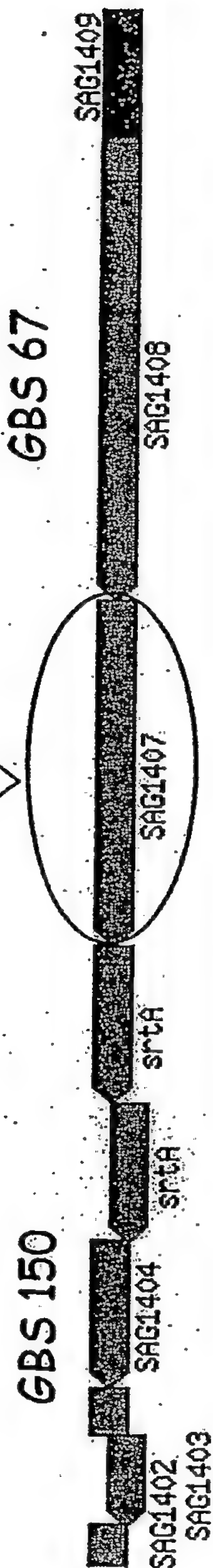


FIGURE 62

Adhesin island 2-
Operon gbs 67, 59, 150



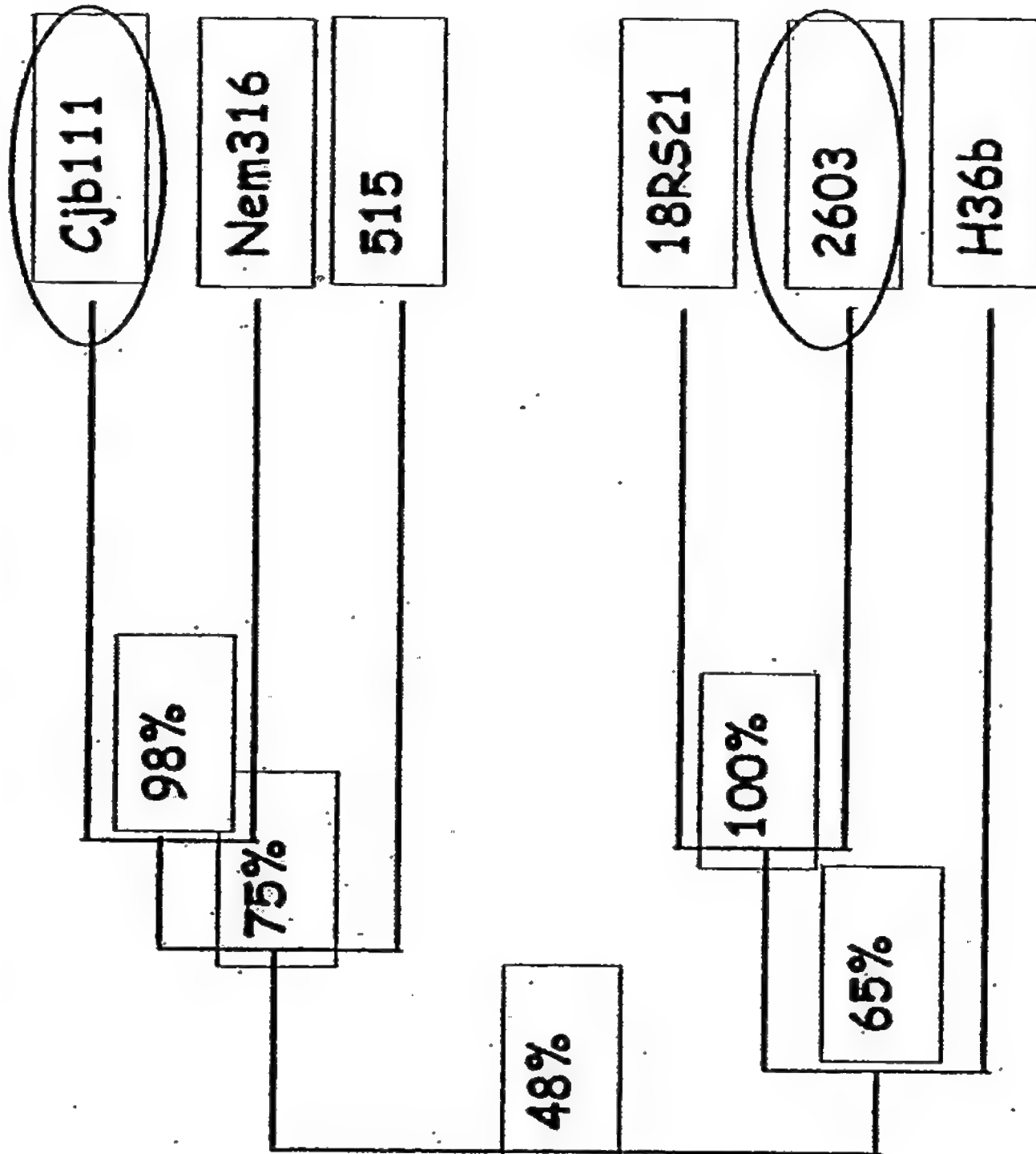
• 30 % identity with GBS 80

• By comparing GBS 59 amino acidic sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

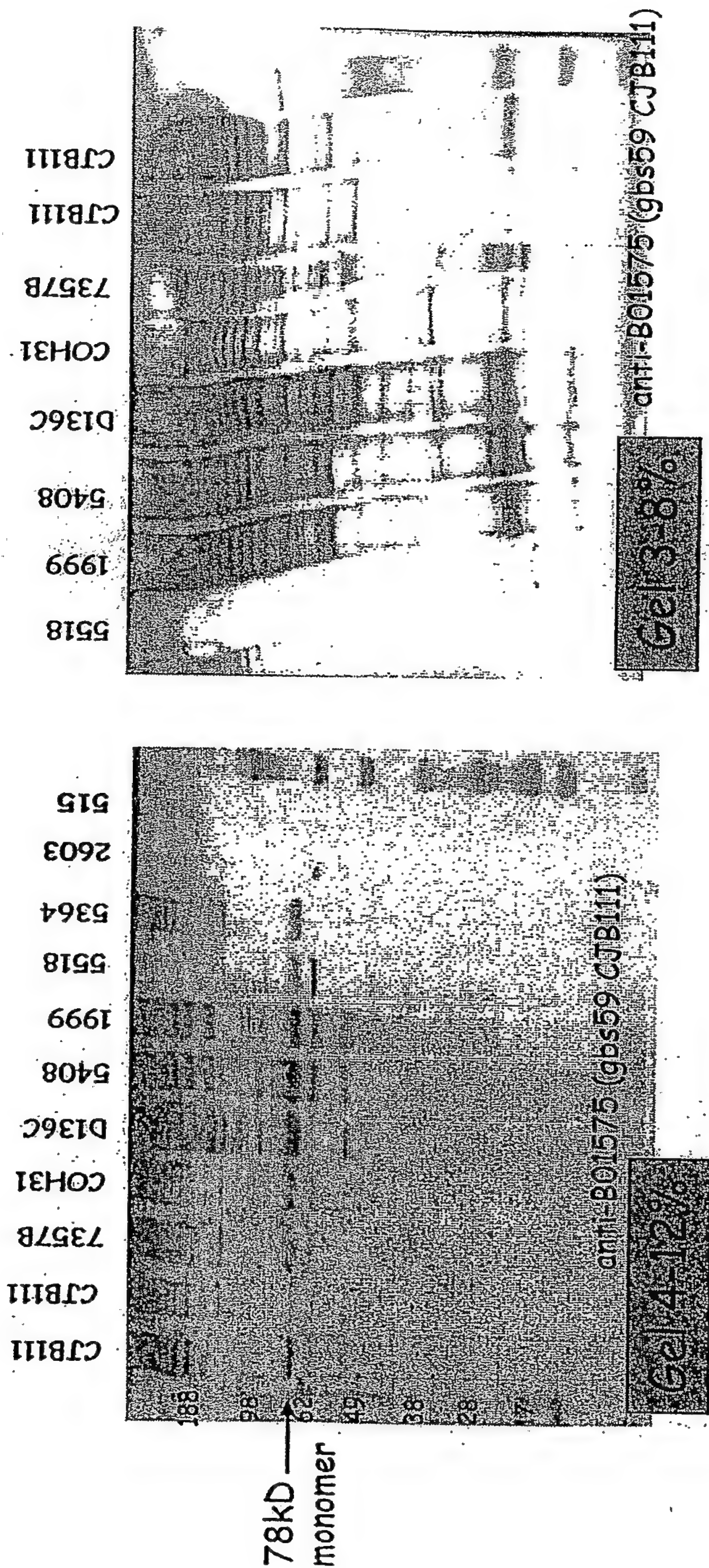
Two-by-two amino acidic sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)
in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,
5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

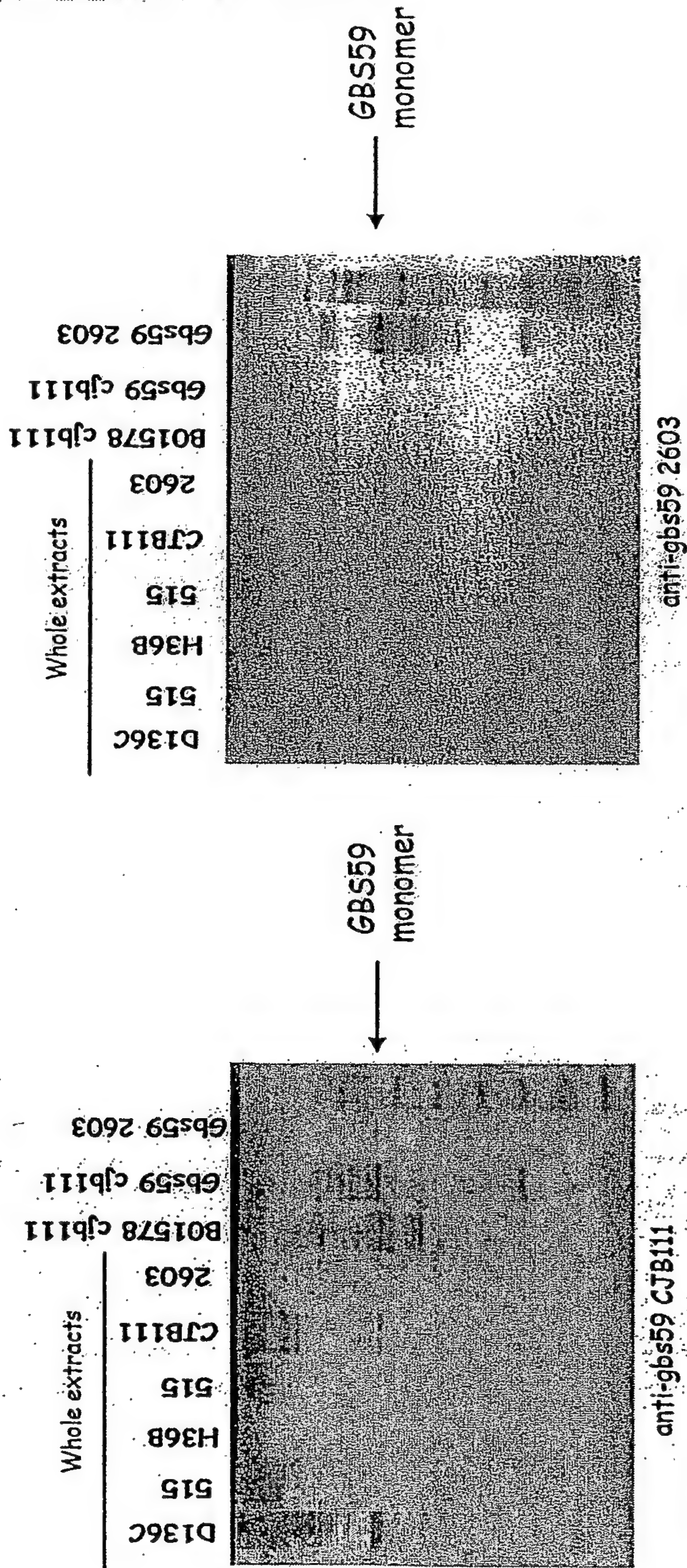


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

WO 2006/078318

216/487

PCT/US2005/027239

PCT/US05/27239

GBS strains	Type	GBS 59
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C	III	504
COH31		505
DK21	II	249
CJB111	V	493
5364		593
2110		590
1999		594
2210		636
5408		537
1169		227

GBS 59

GBS 67

PBS

CJB111

7357B

515

Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66

Opsonophagocytosis assays: B01575 (gbs59-cjb111)

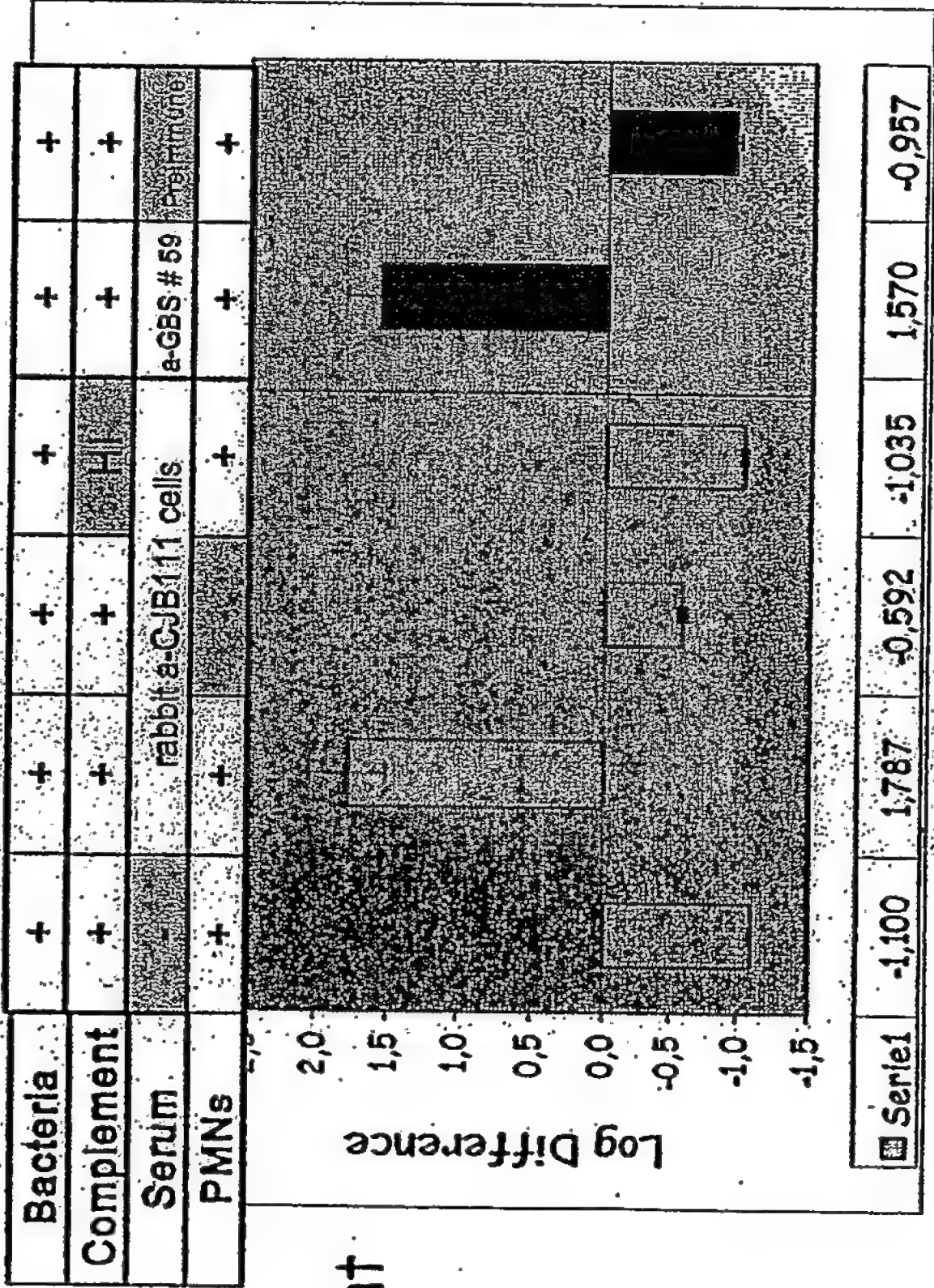
WO 2006/078318

PCT/US05/27239

217/487

PCT/US2005/027239

I
experiment



- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V

II
experiment

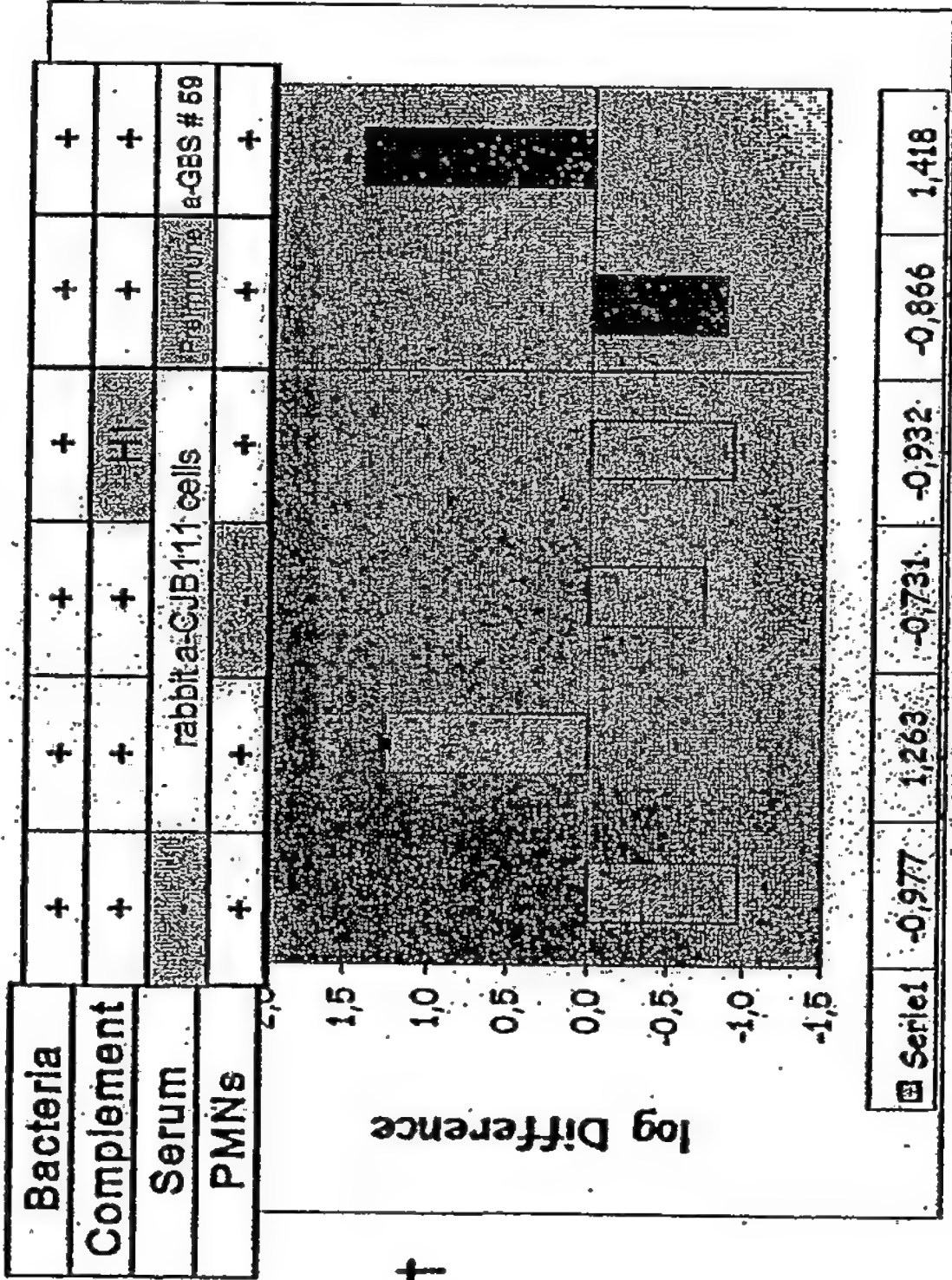


FIGURE 67

Association GBS 80-104 WB α -80, α -104 JM9130013 Total Extract

Mab α -80 A477
Mab α -80 19G4178
Mab α -104 15H3149
Mab α -104 12A767
Mab α -104 H2132
Mab α -104 14F3173
 α -104 POLIC.
 α -80 POLIC.
Mab α -80 19G4177

FACS	(Δ MEAN)
GBS 80	597
GBS 104	446

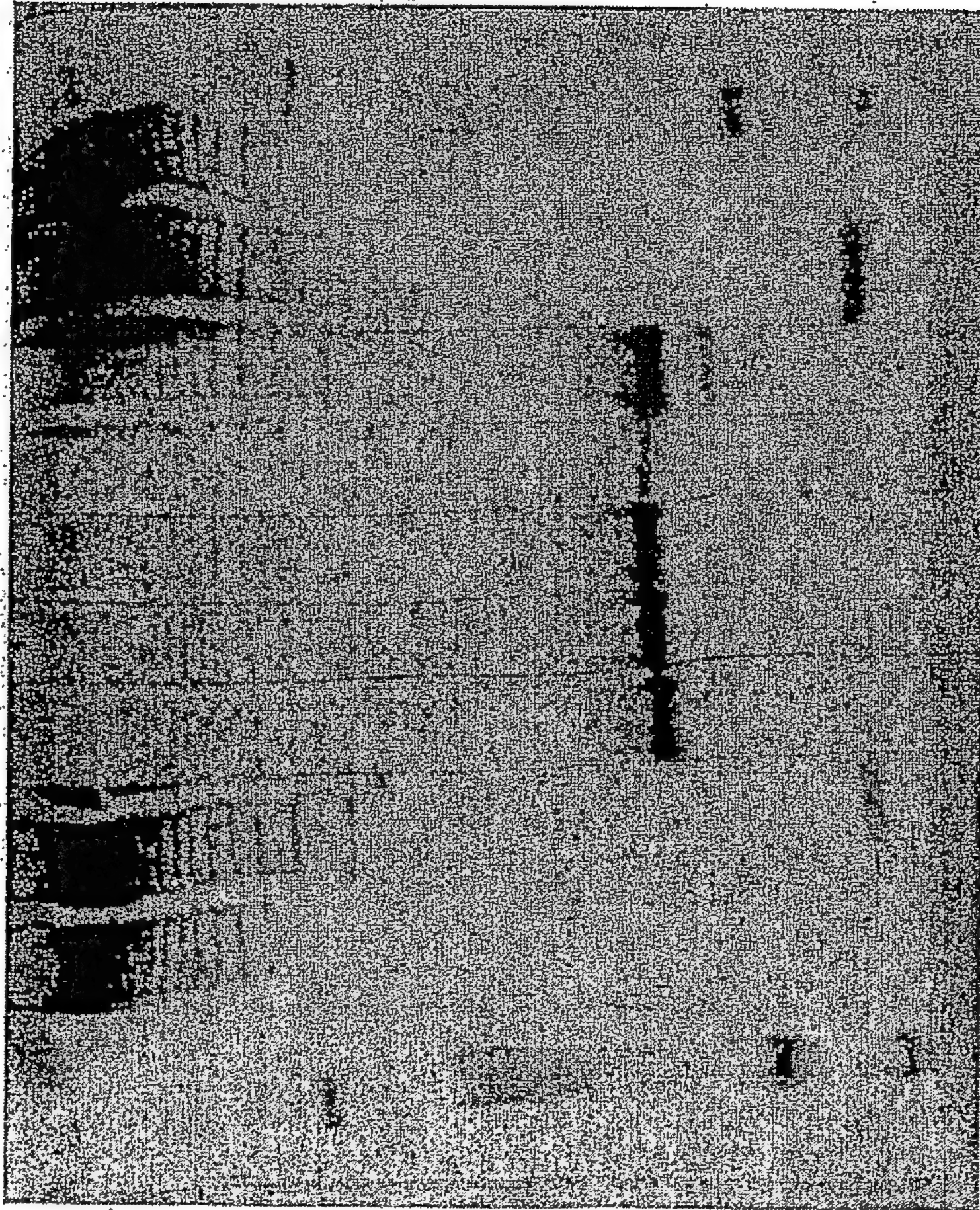
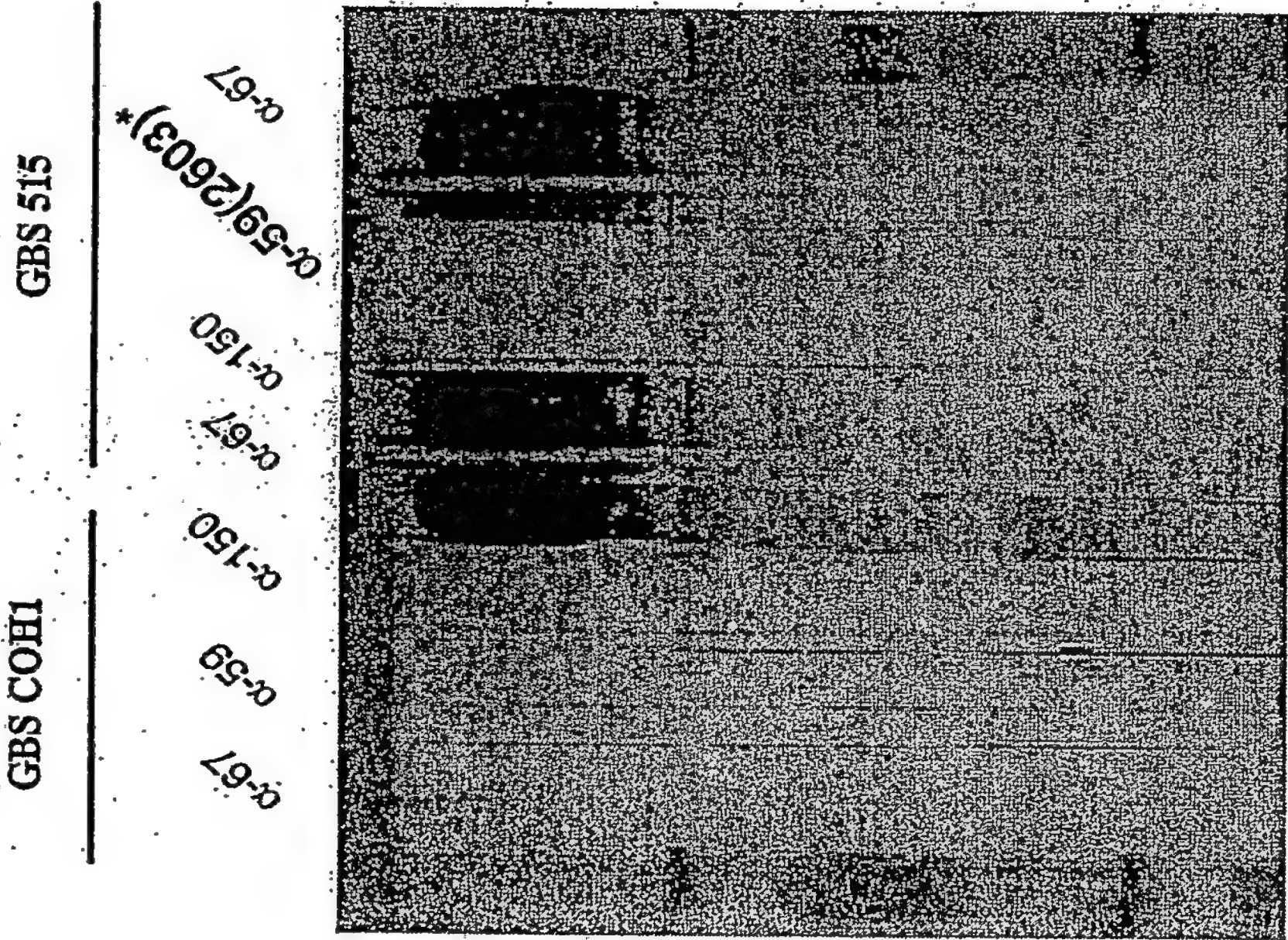


FIGURE 68

WB GBS 515 Total Extract
 α -67; α -150; α -59 (2603)



Controls:
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of
a high molecular weight polymer (pilus)
in 515 GBS strain

FIGURE 69

Western Blotting ko GBS67 from 515 genome (clone 1.45)

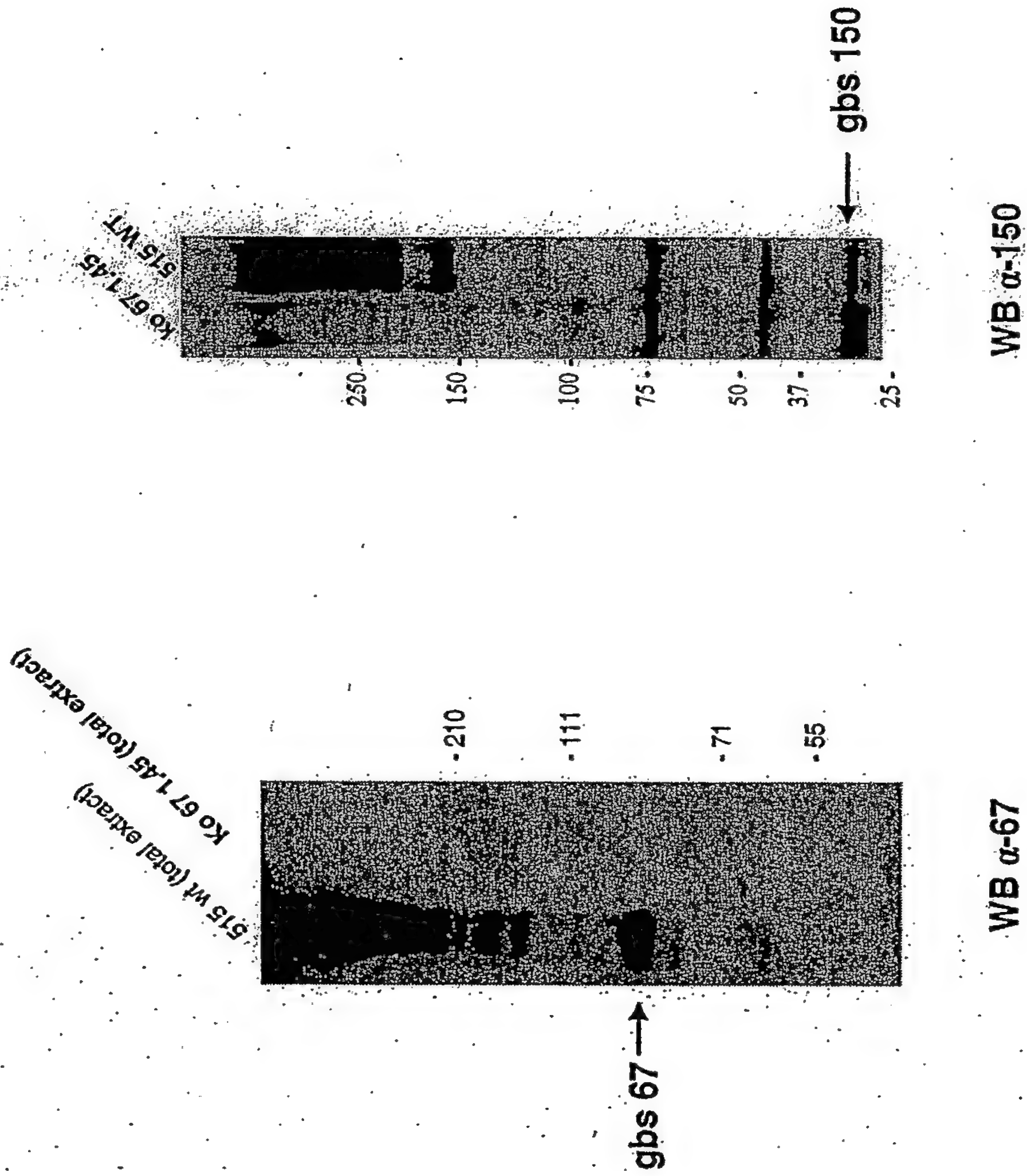
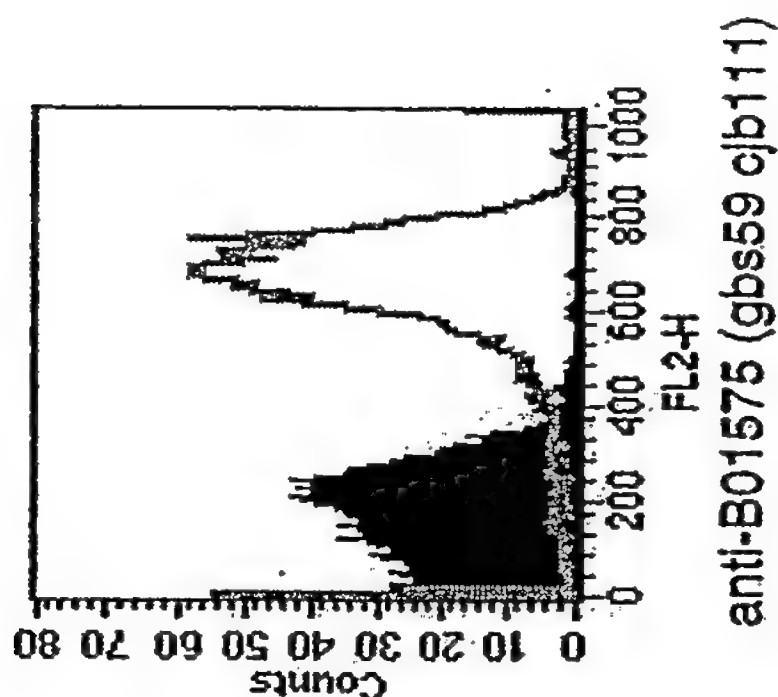
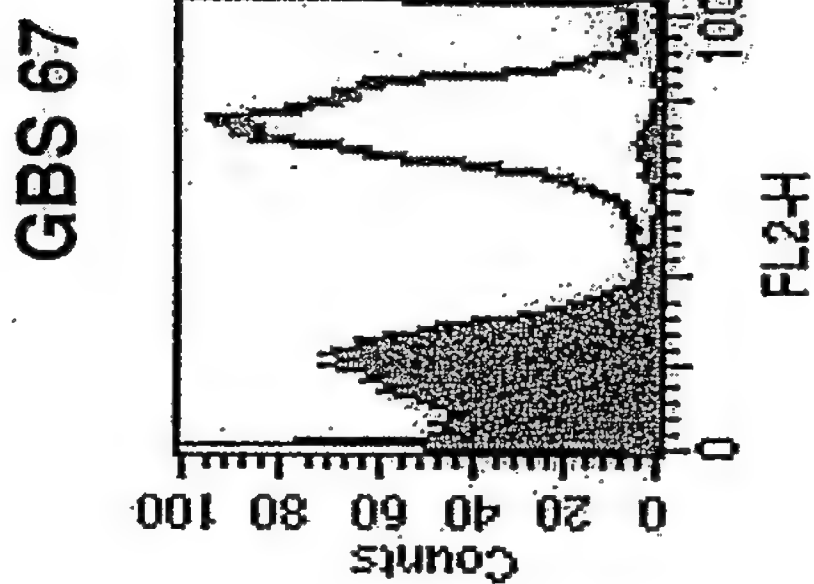
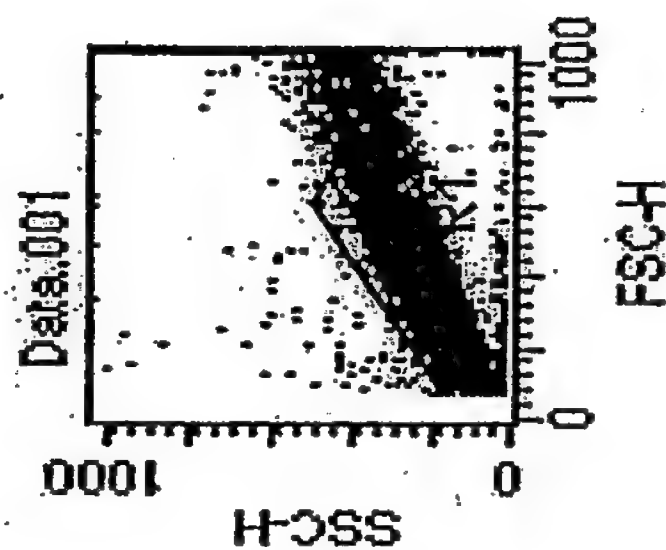


FIGURE 70

FACS GBS 515 Δ67

515 WT



515 Δ67
clone 1.45

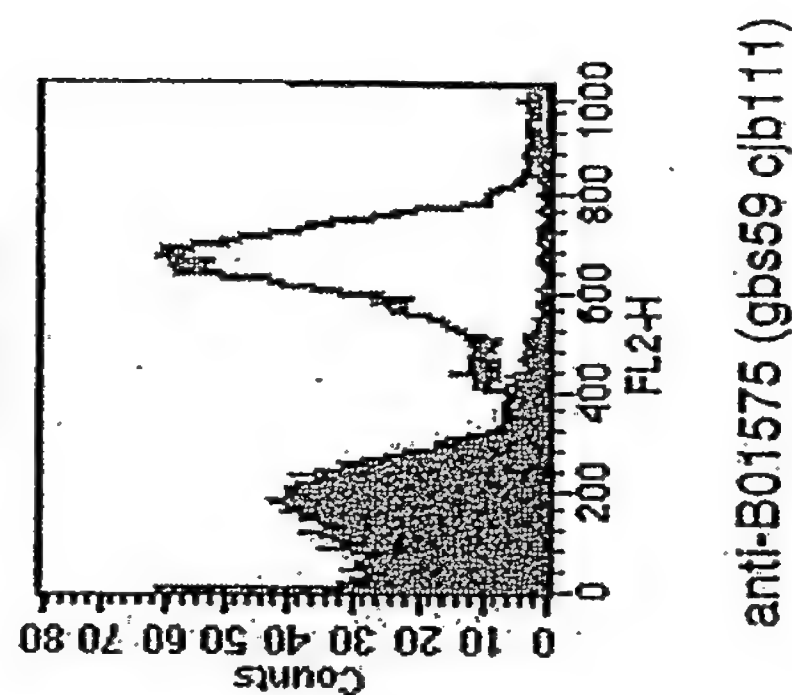
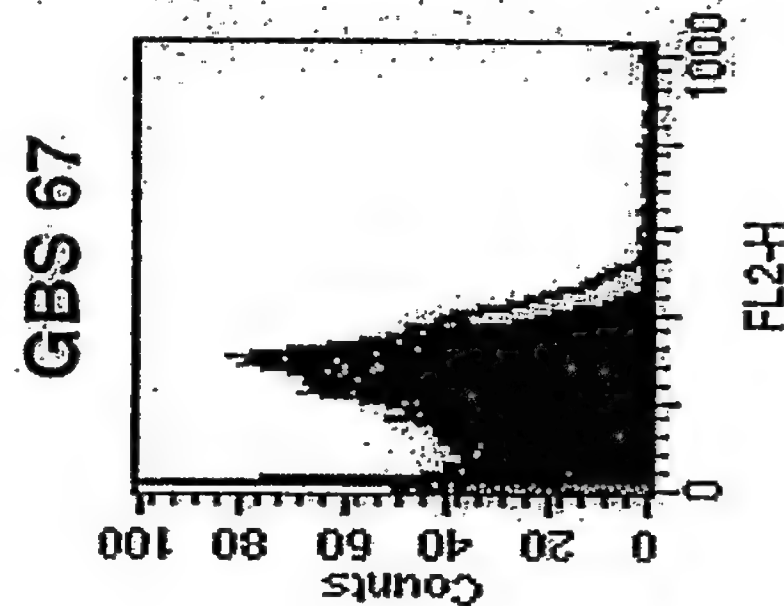
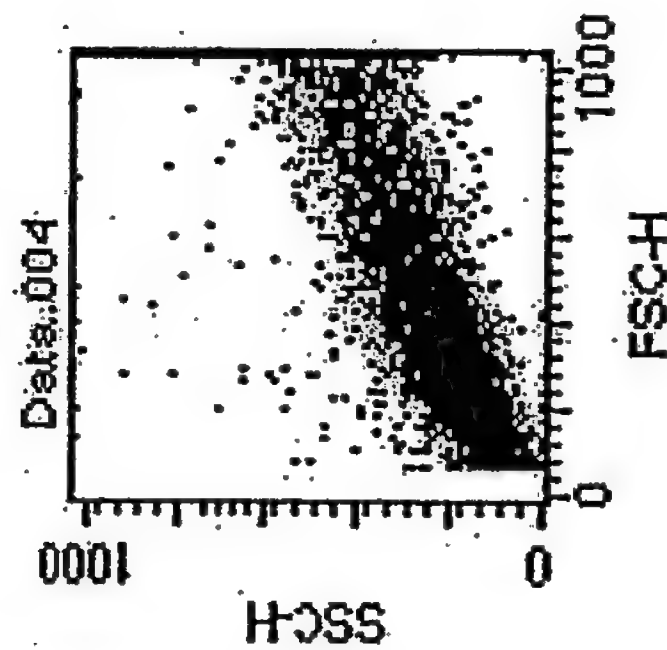
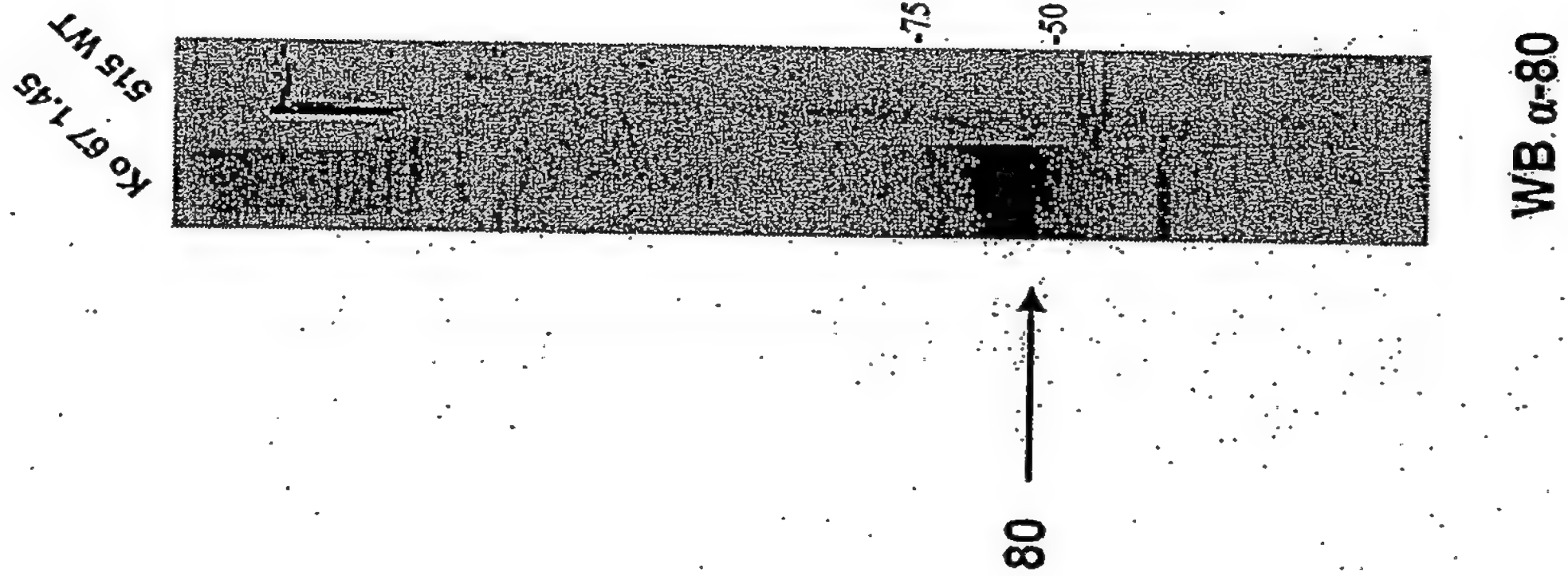


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80



GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

FIGURE 72

spyM6_0159 type 1 pilus present in M6

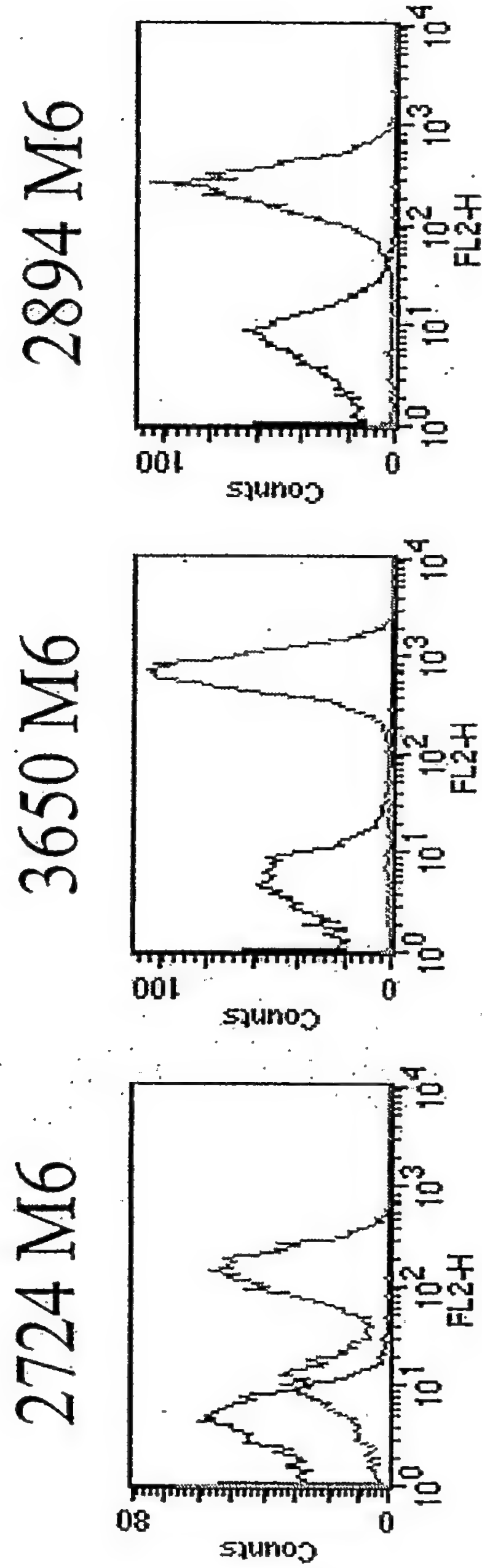


Figure 73

PCT/US05/27239

224/487

spyM6_0160 type 1 pilus present in M6

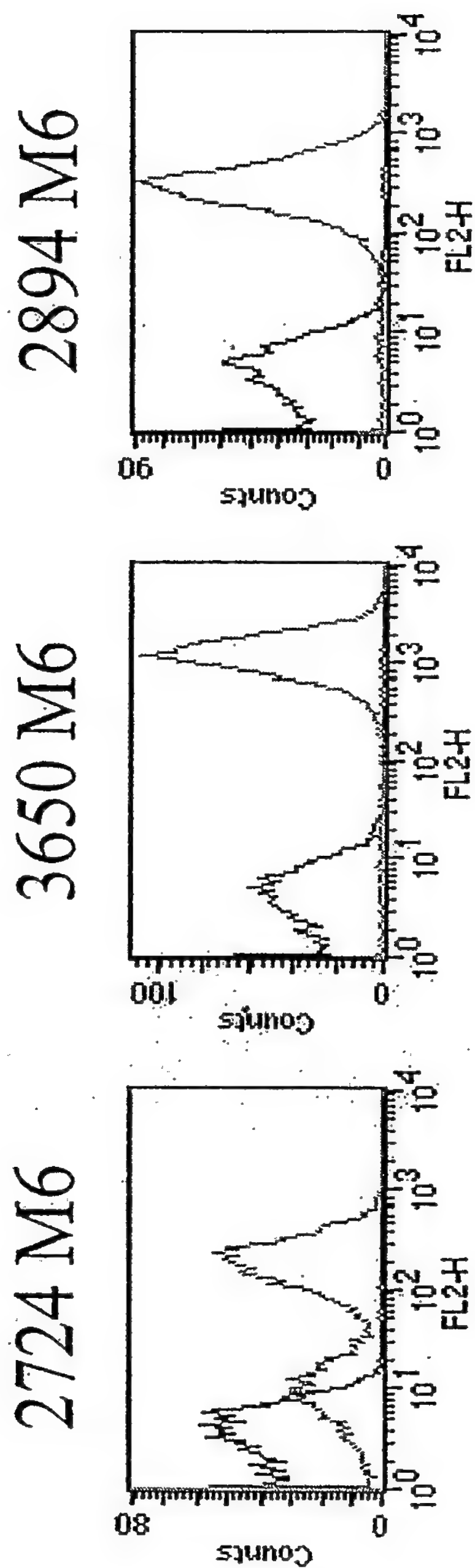


Figure 74

225/487

Gas15 type 2 pilus present in M1

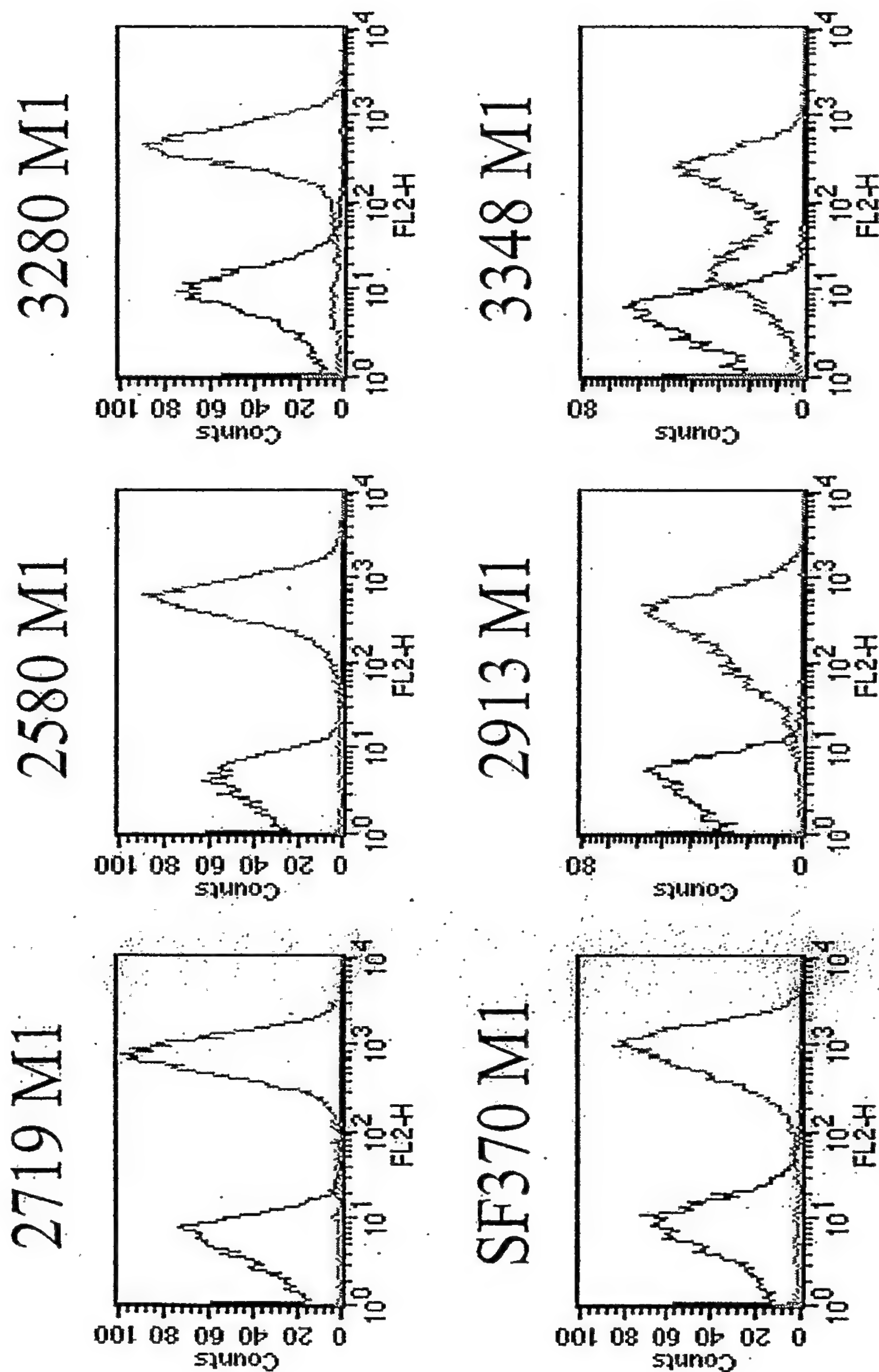


Figure 75

Gas16 type 2 pilus present in M1

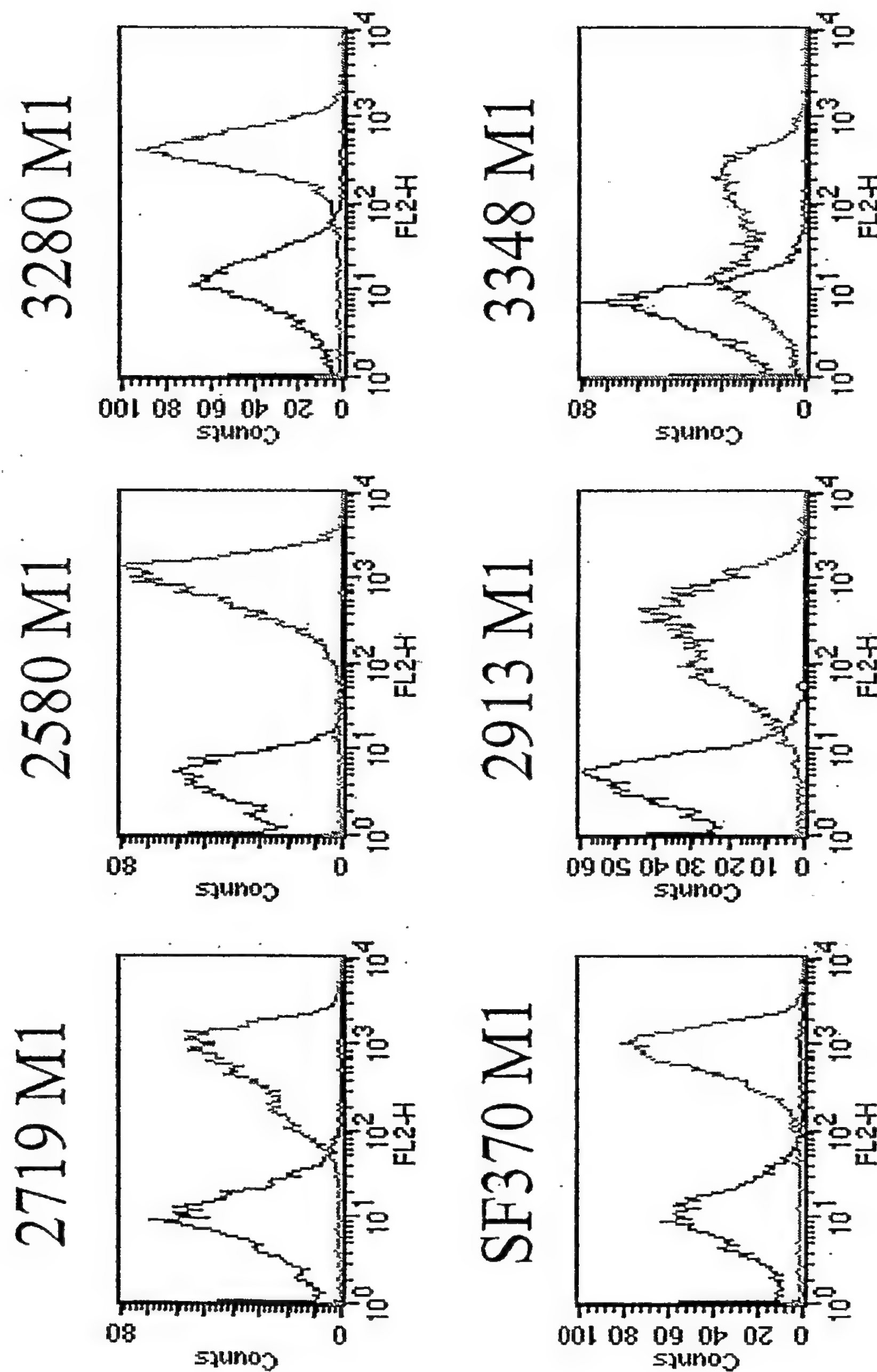


Figure 76

Gas18 serum 1 type 2 pilus present in M1

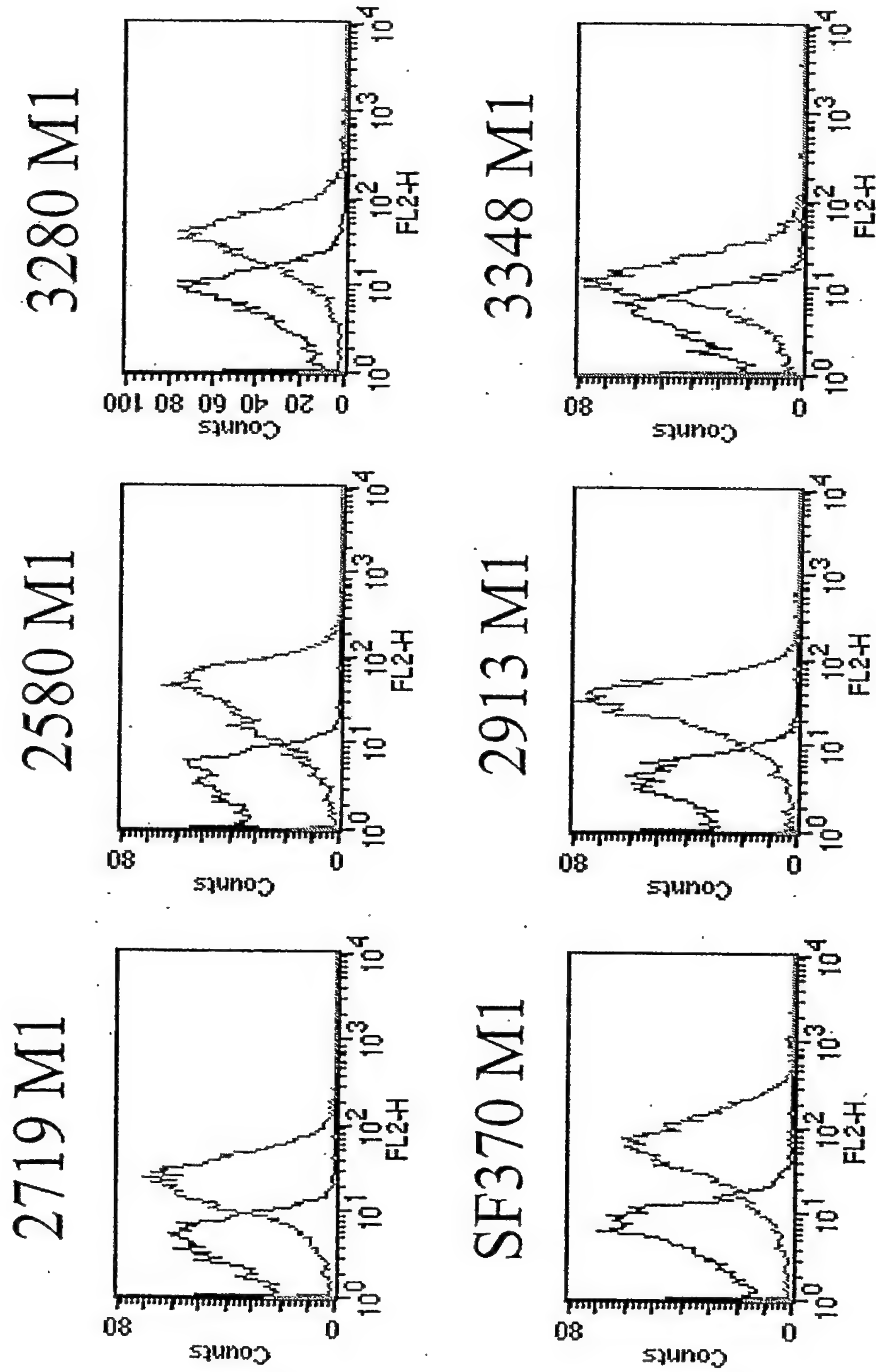


Figure 77

Gas18 serum 2 type 2 pilus present in M1

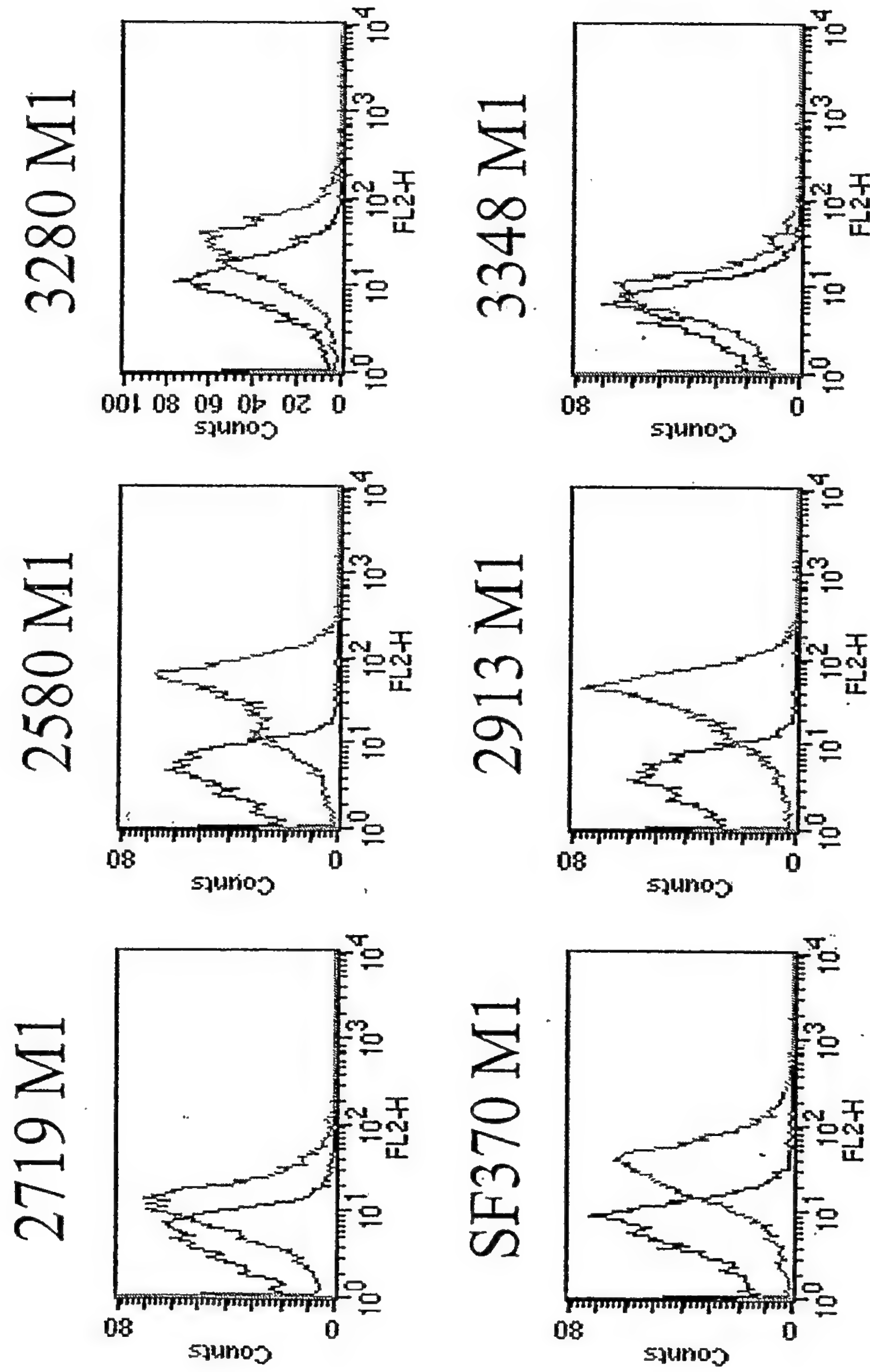


Figure 78

229/487

Gas16p2 type 2 pilus present in M1

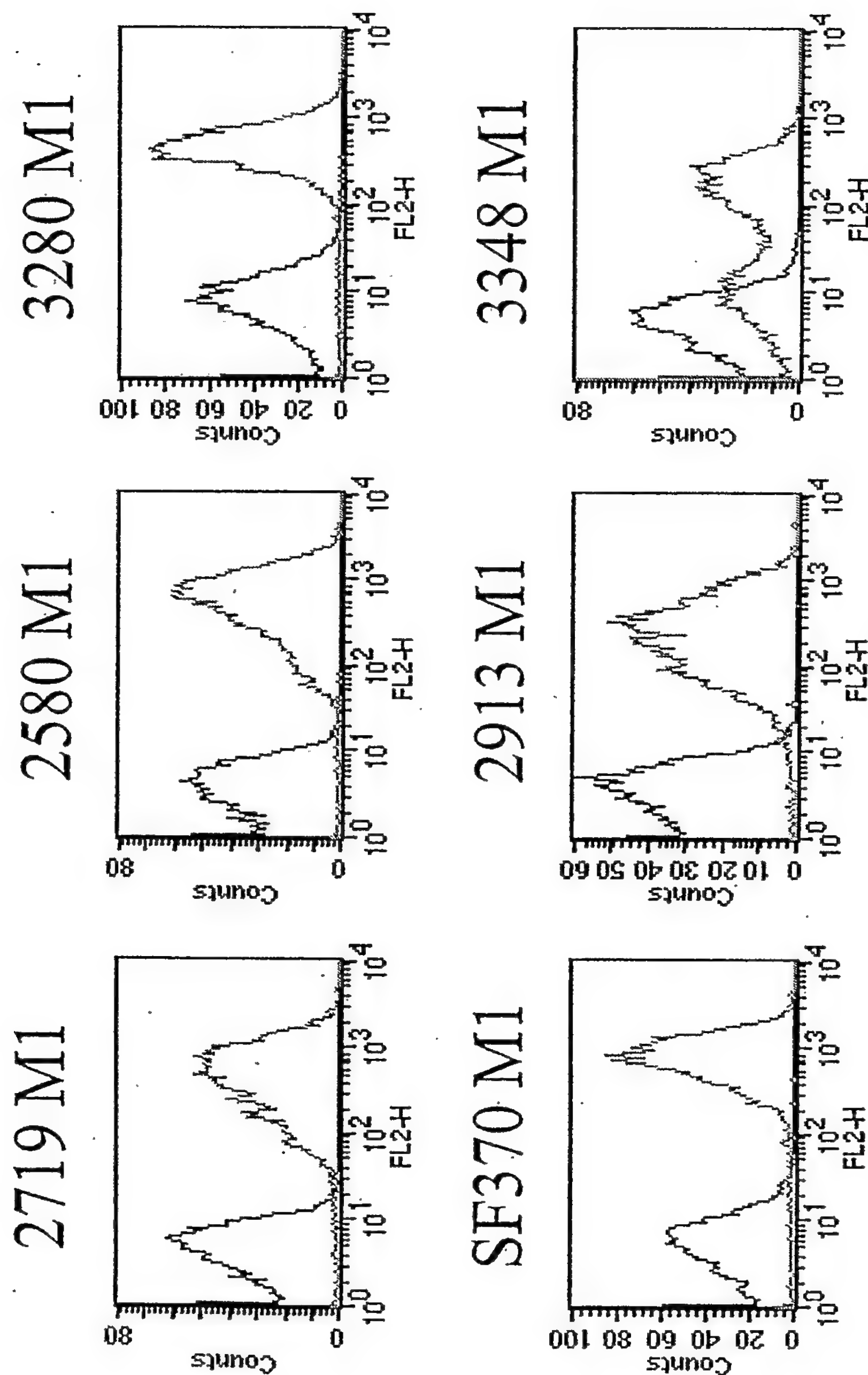


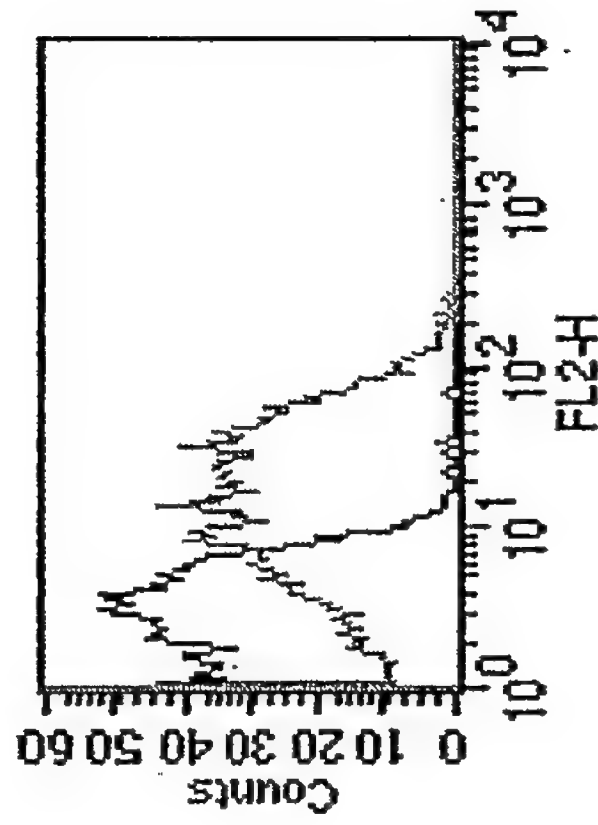
Figure 79

PCT/US05/27239

230/487

spyM3_0098 type 3 pilus present in M3

3135 M3



2721 M3

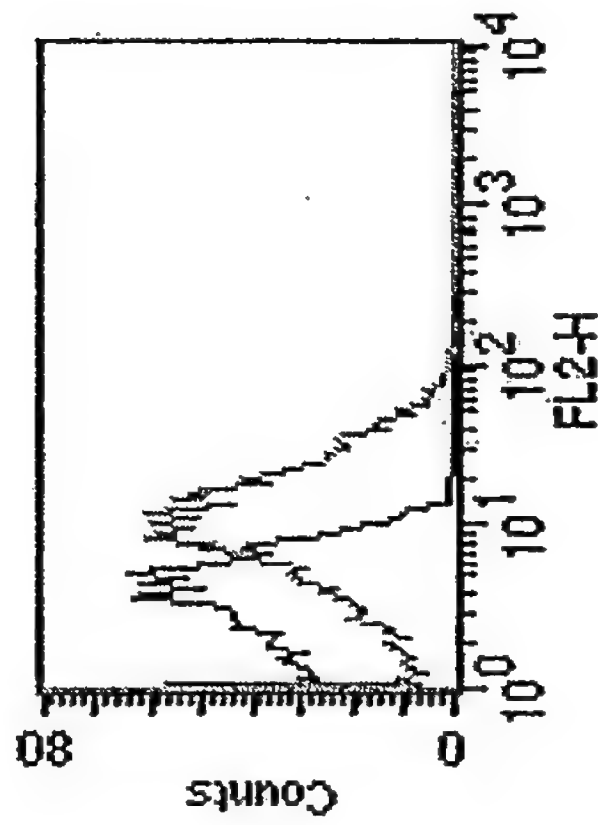
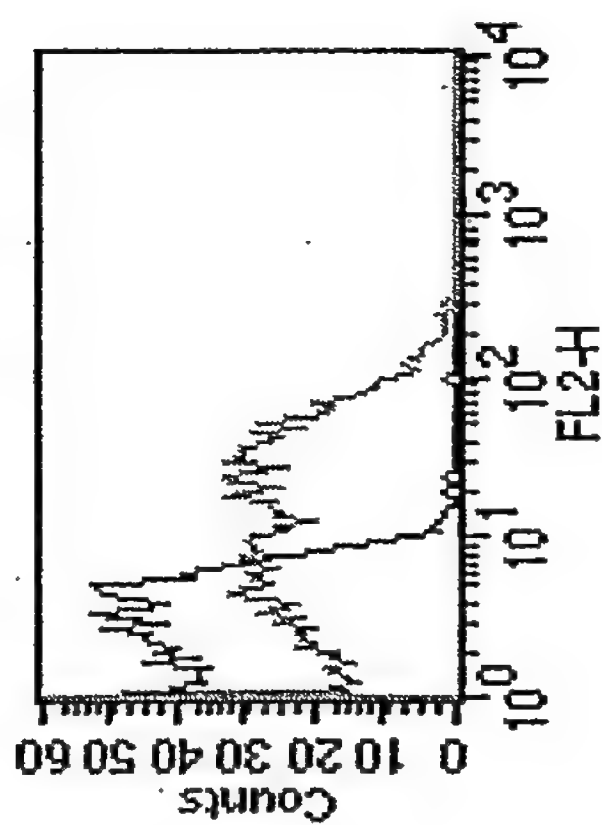


Figure 80

spyM3_0100 type 3 pilus present in M3

3135 M3



2721 M3

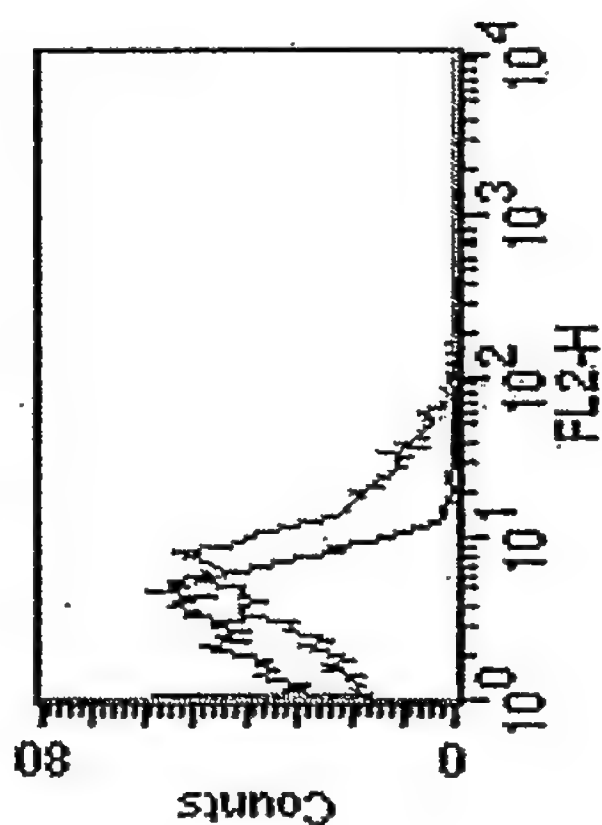
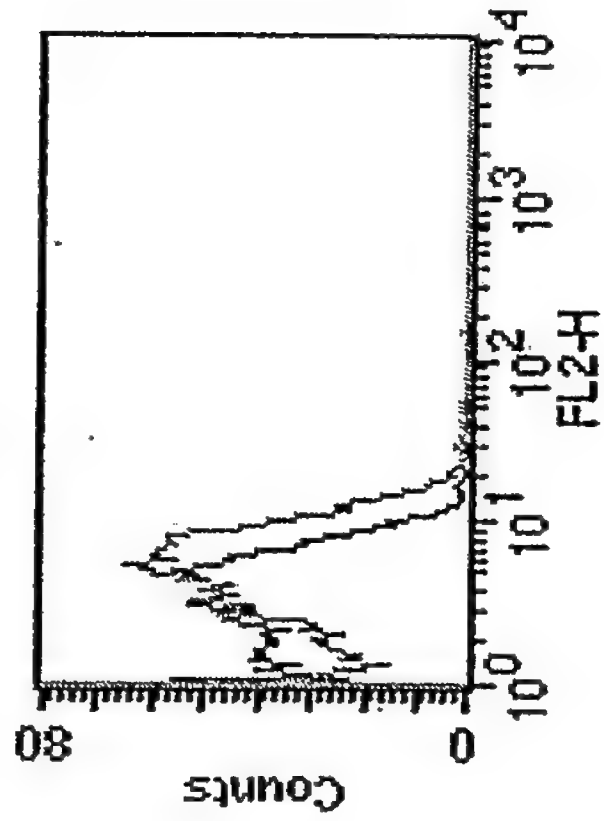


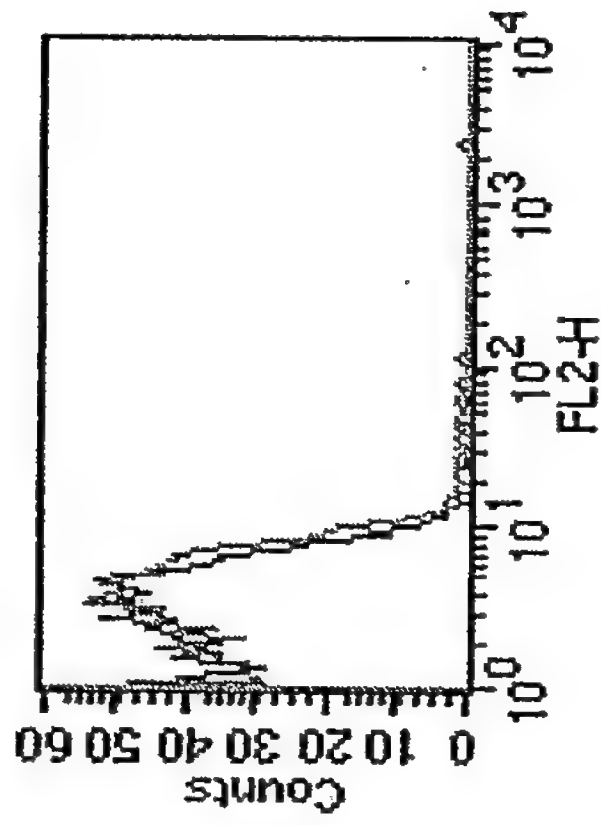
Figure 81

spyM3_0102 type 3 pilus present in M3

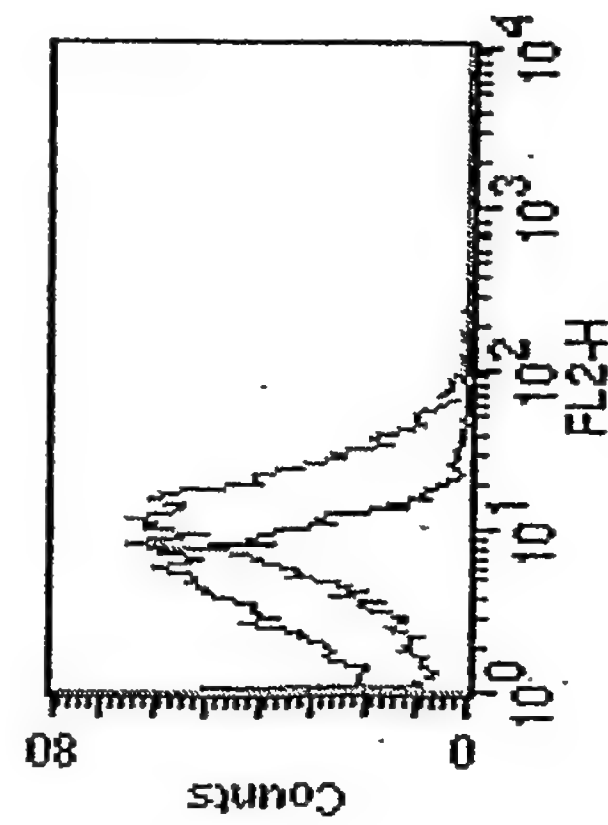
2721 M3



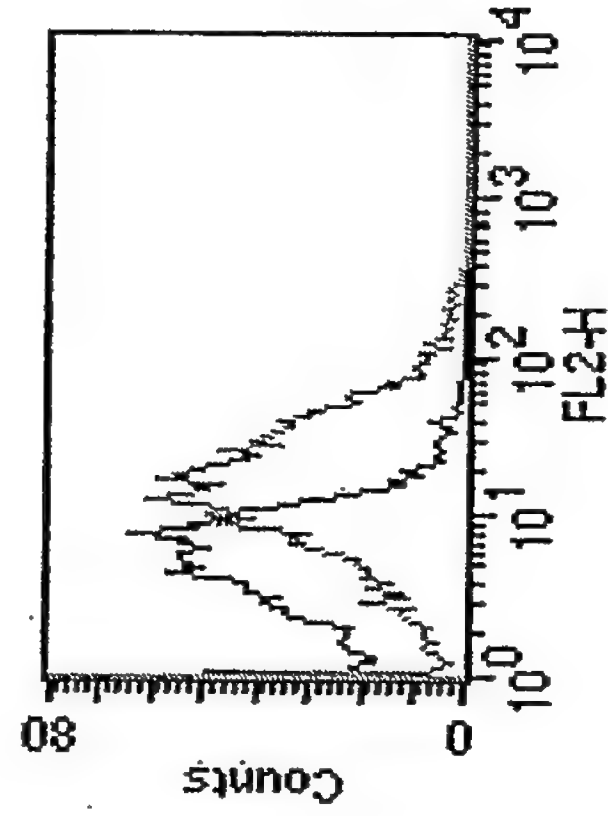
3135 M3



2724 M6



3650 M6



2894 M6

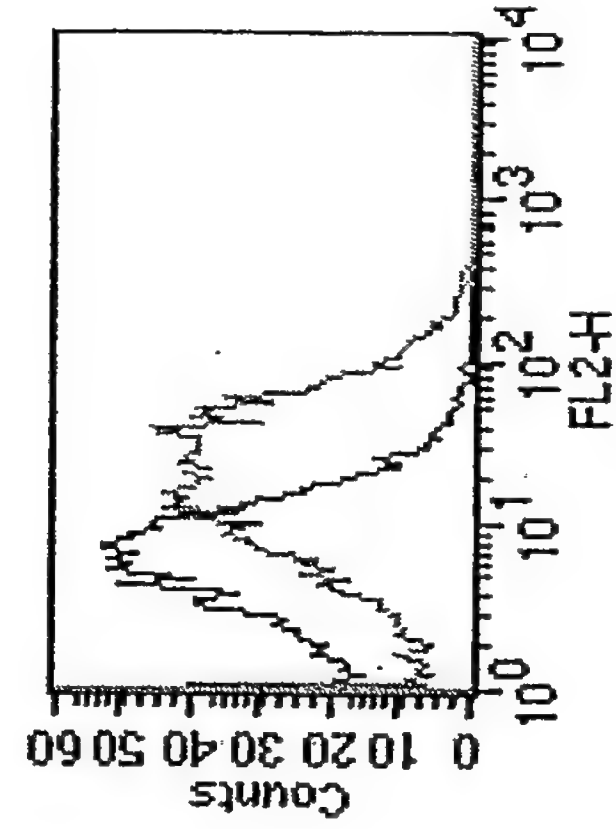
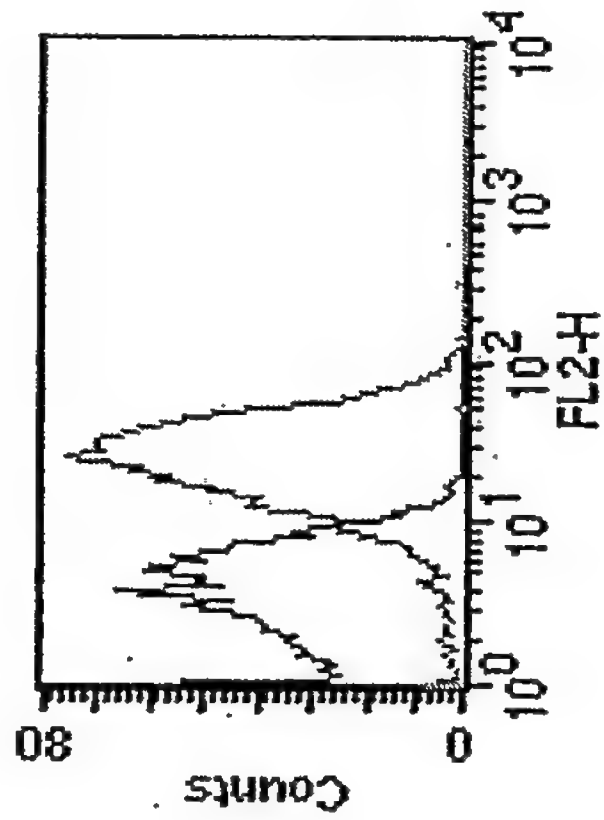


Figure 82

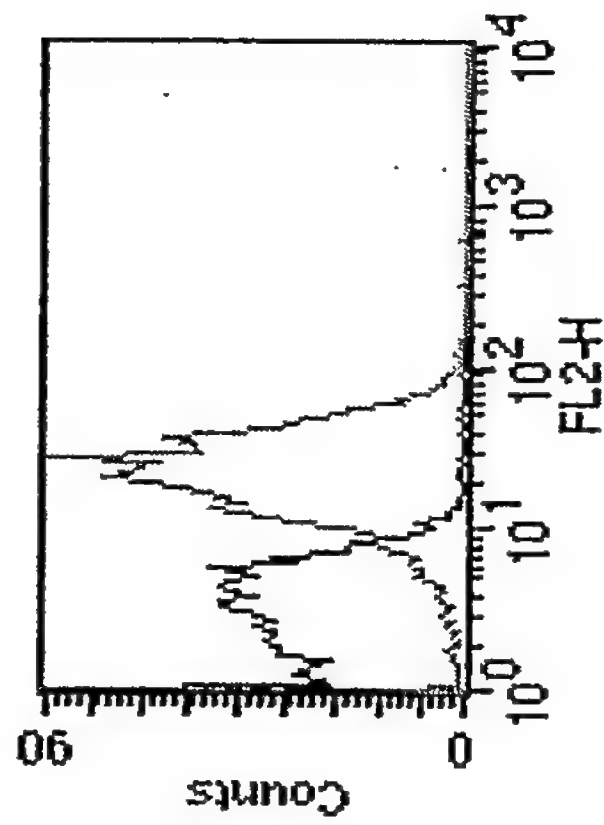
PCT/US05/27239 233/487

spyM3_0104 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

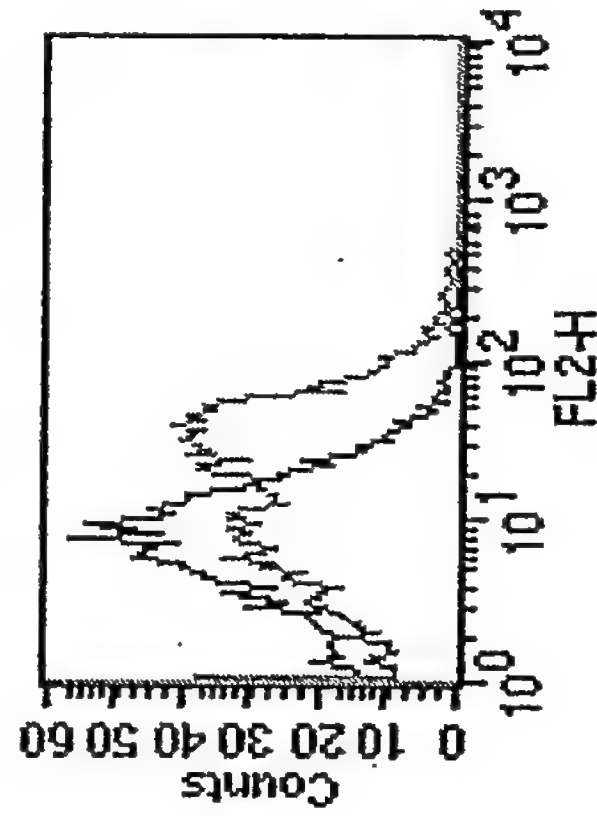
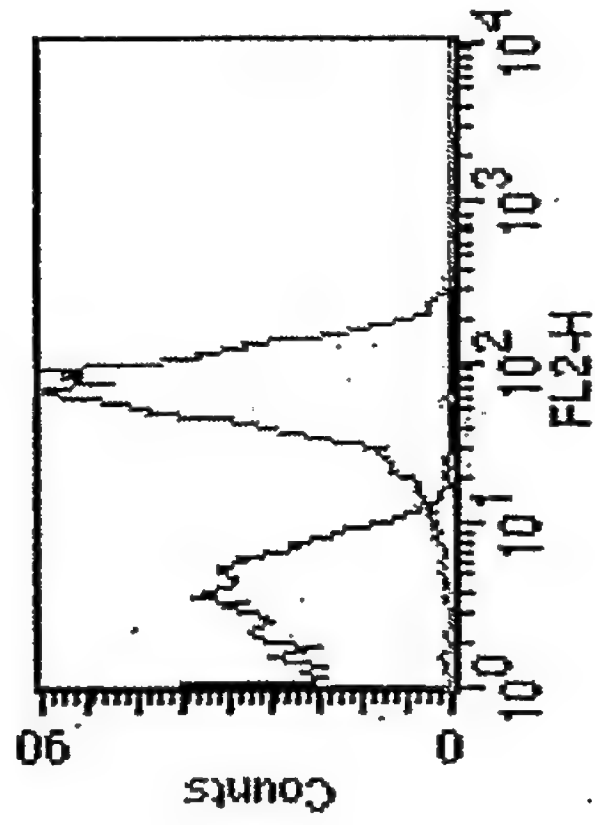


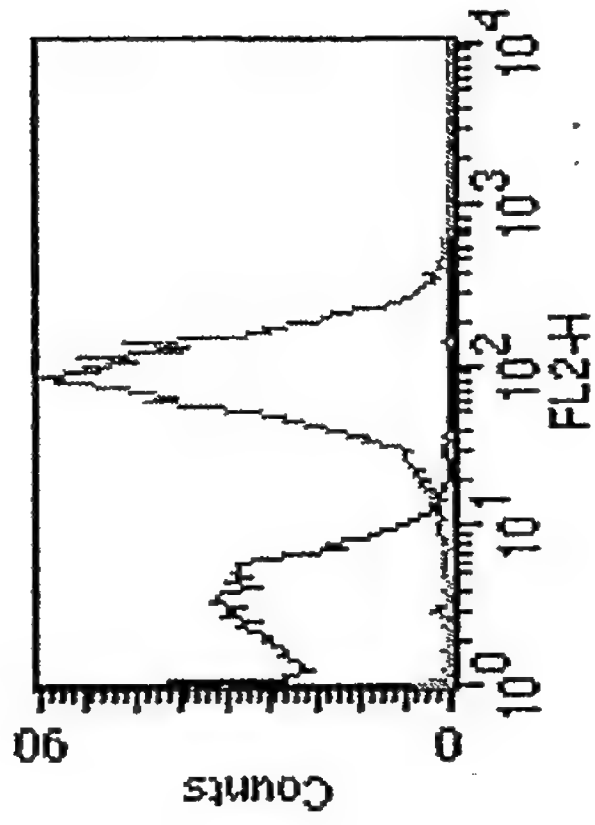
Figure 83

spyM3_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

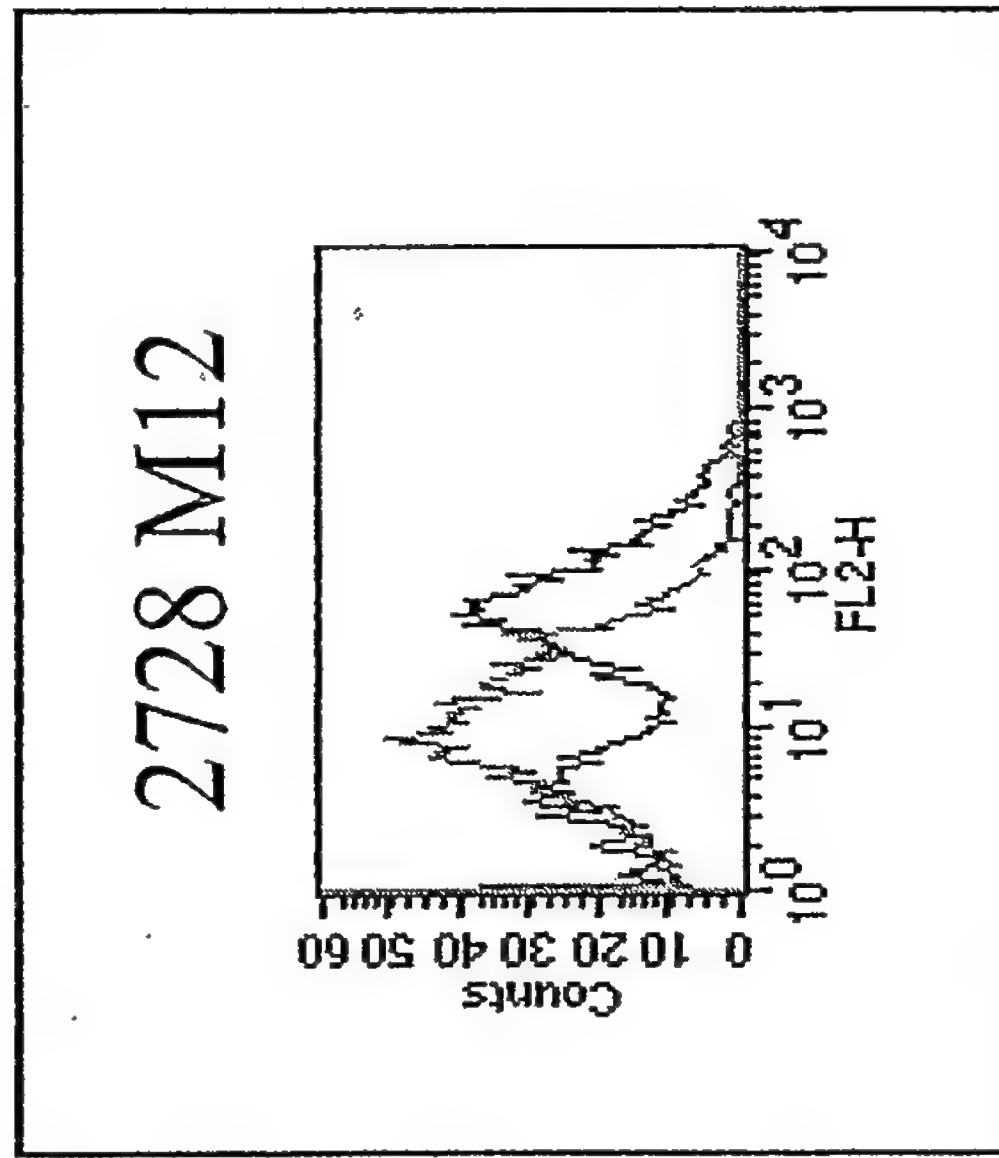
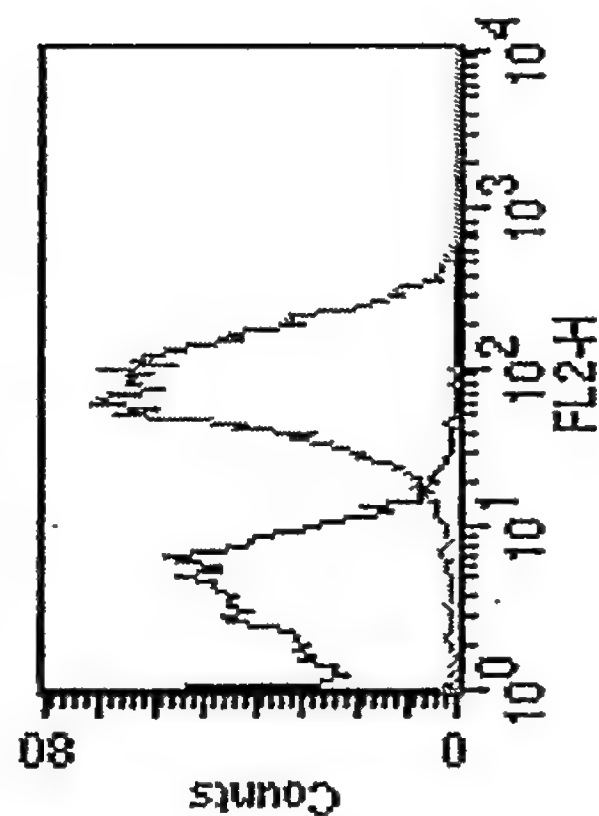


Figure 84

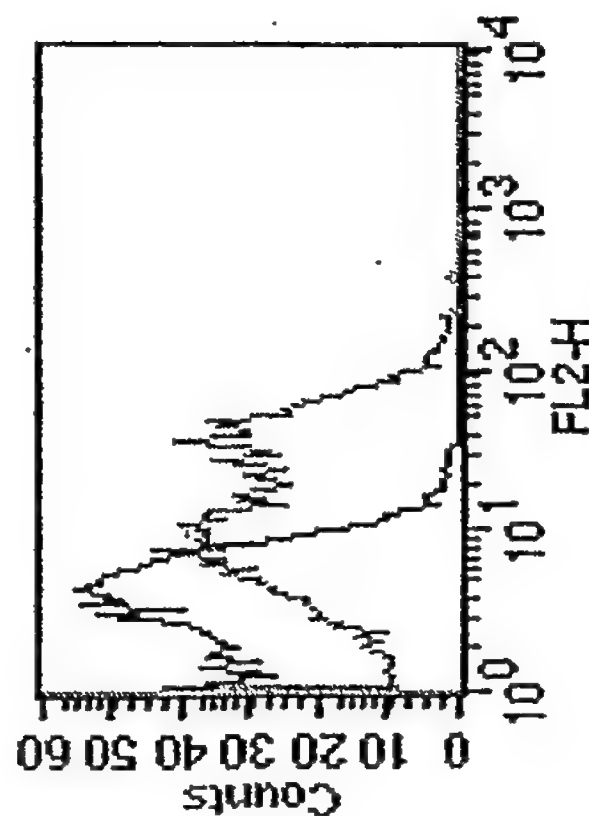
PCT/US05/27239 235/487

19224134 type 4 pilus present in M12

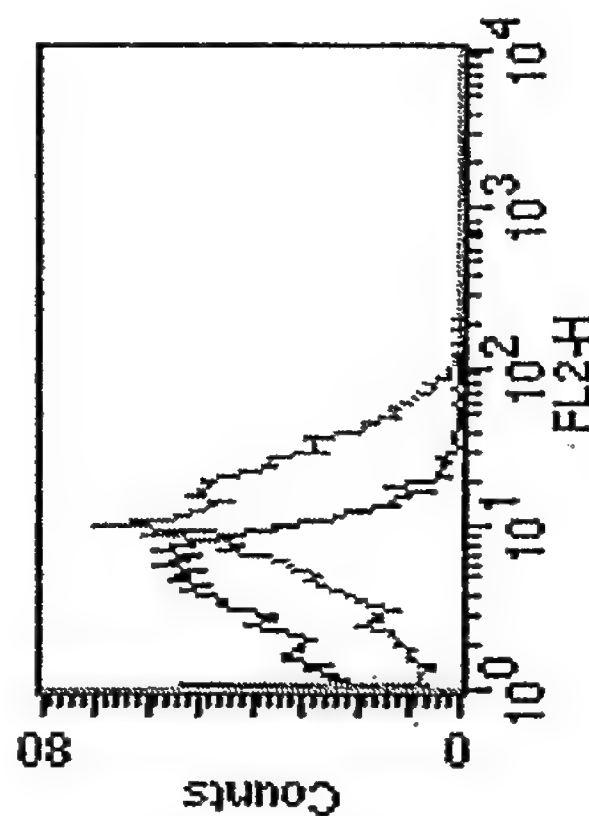
2728 M12



2724 M6



3650 M6



2894 M6

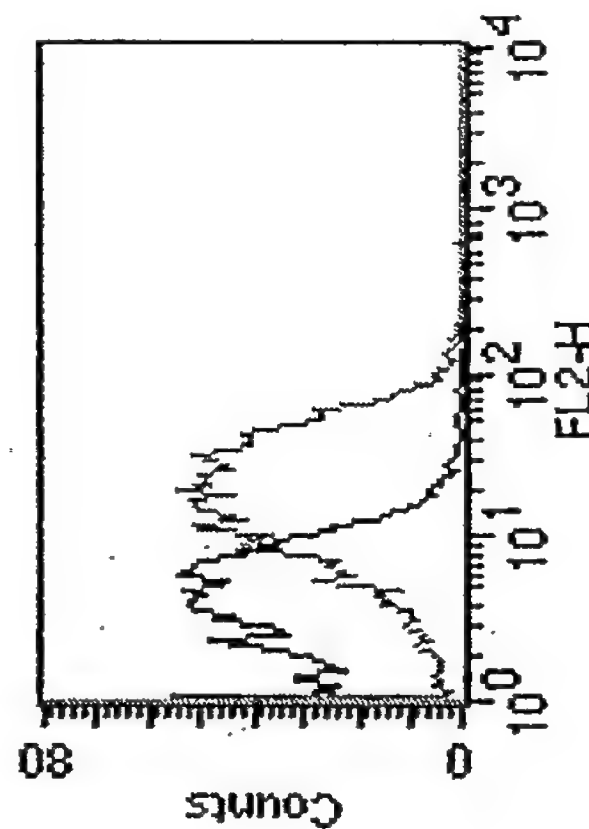


Figure 85

19224135 type 4 pilus present in M12

2728 M12

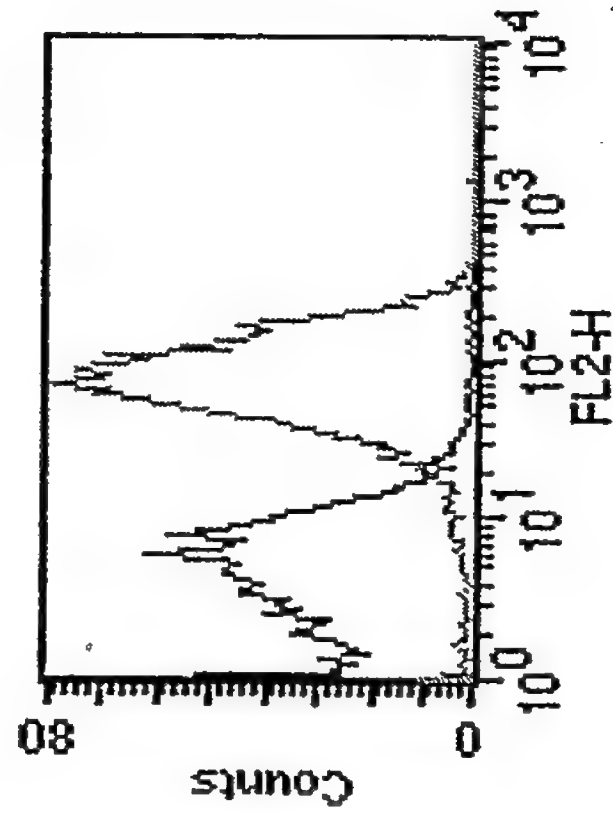


Figure 86

PCT/US05/27239

237/487

19224137 type 4 pilus present in M12

2728 M12

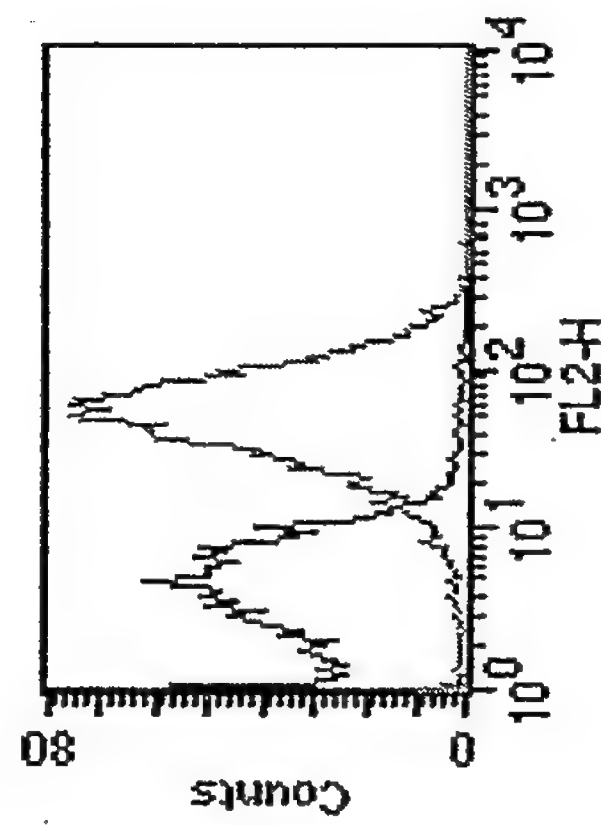


Figure 87

19224141 type 4 pilus present in M12

2728 M12

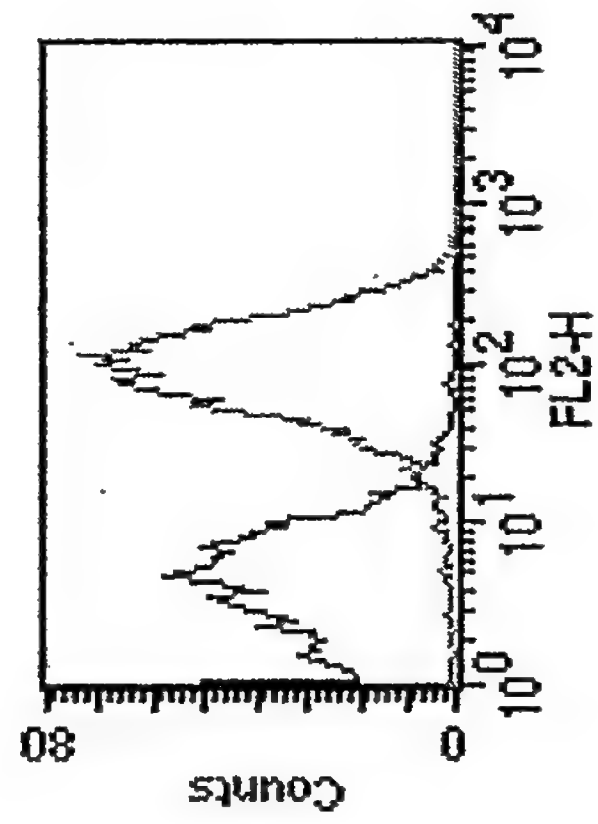


Figure 88

Figure 89

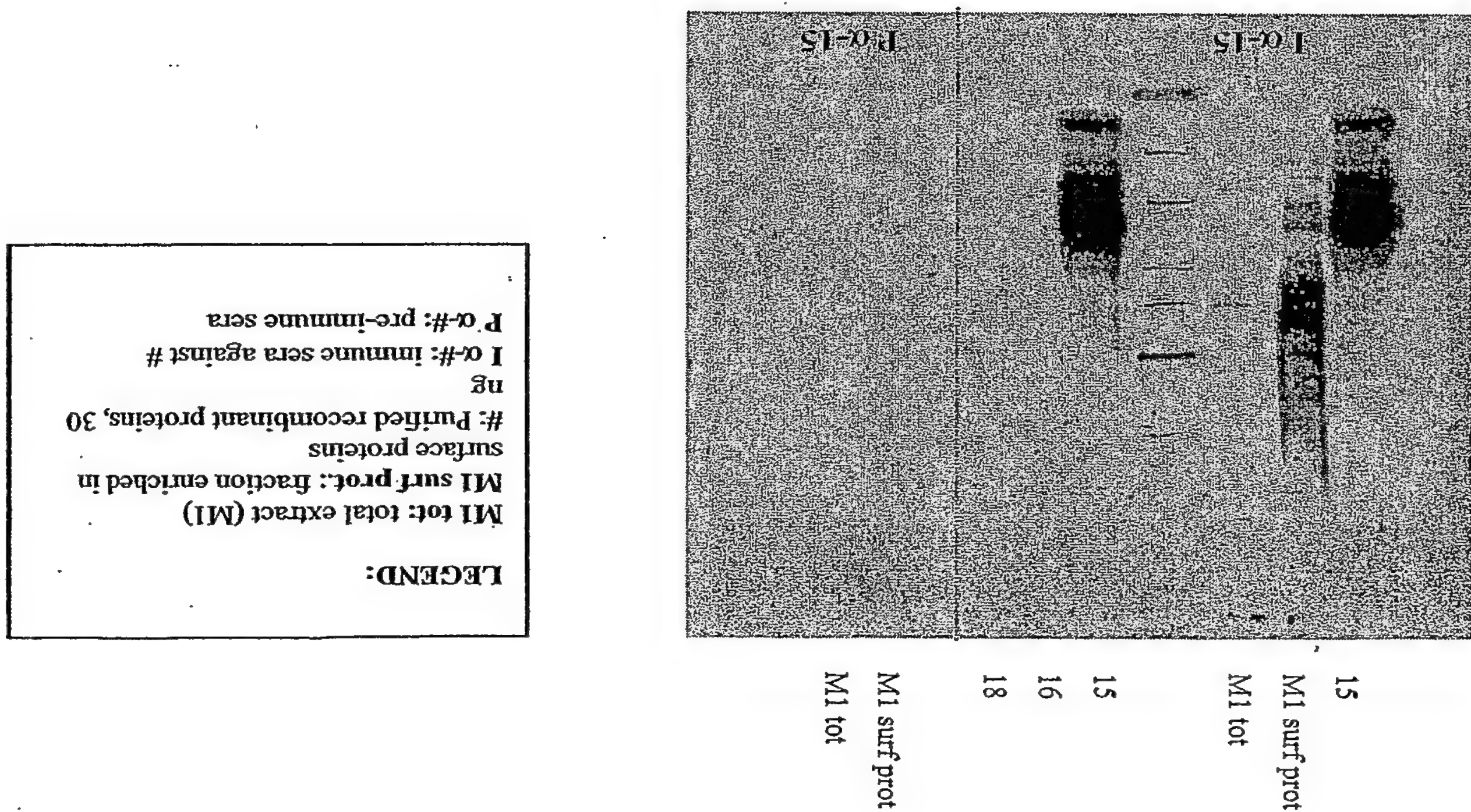
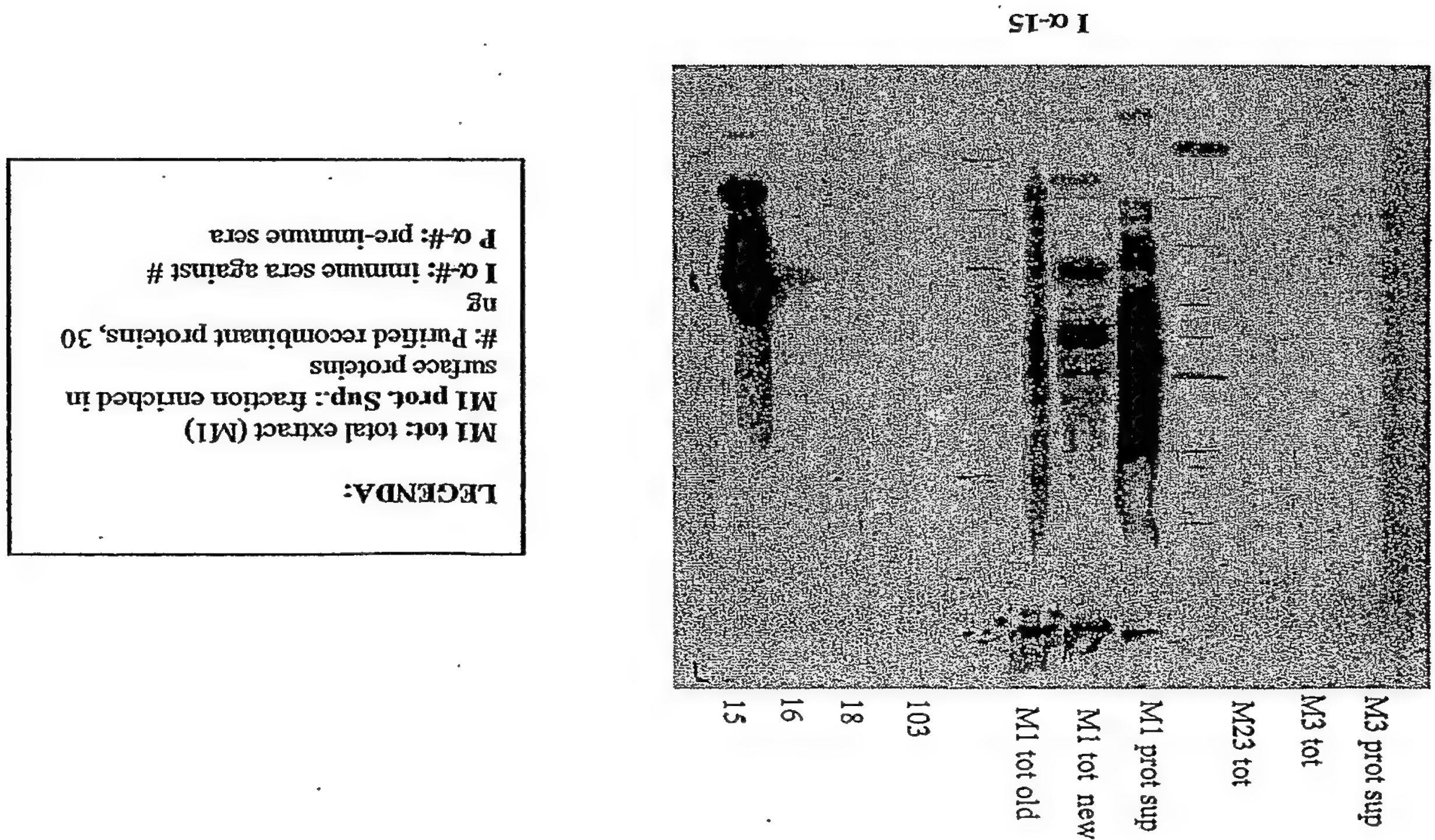


Figure 90



PCT/US05/27239 241/487

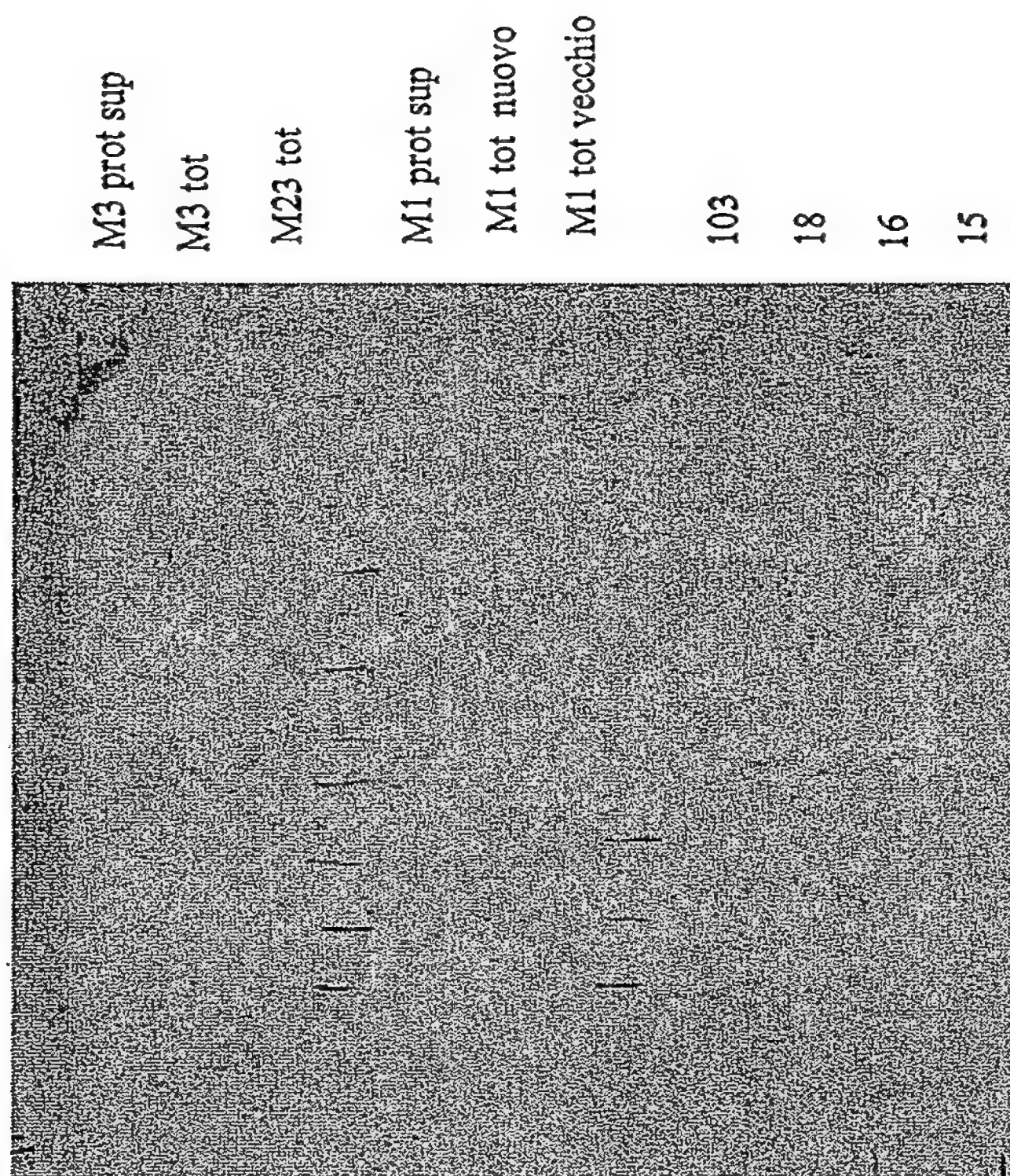
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-15**

Figure 91

PCT/US05/27239

242/487

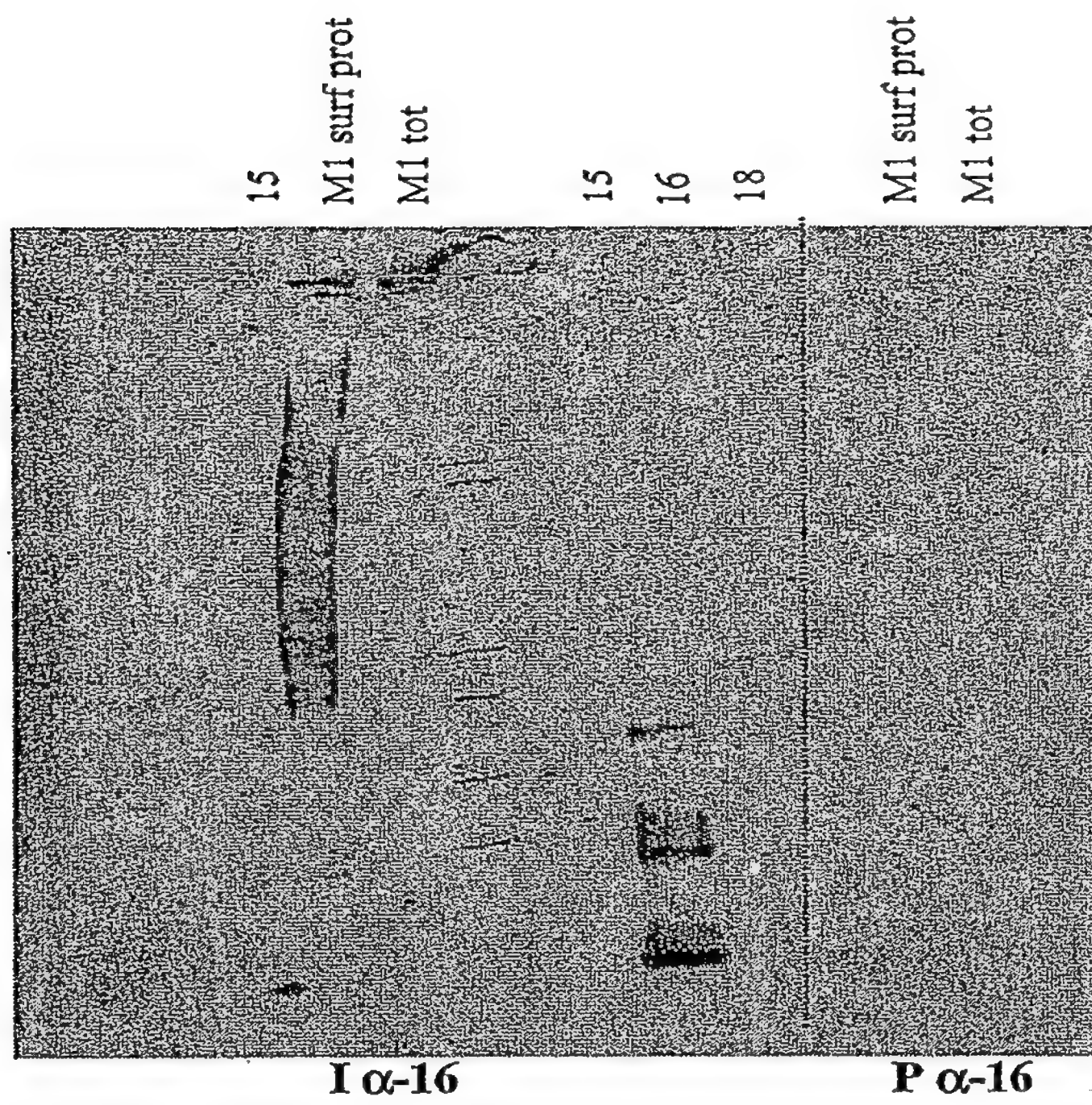
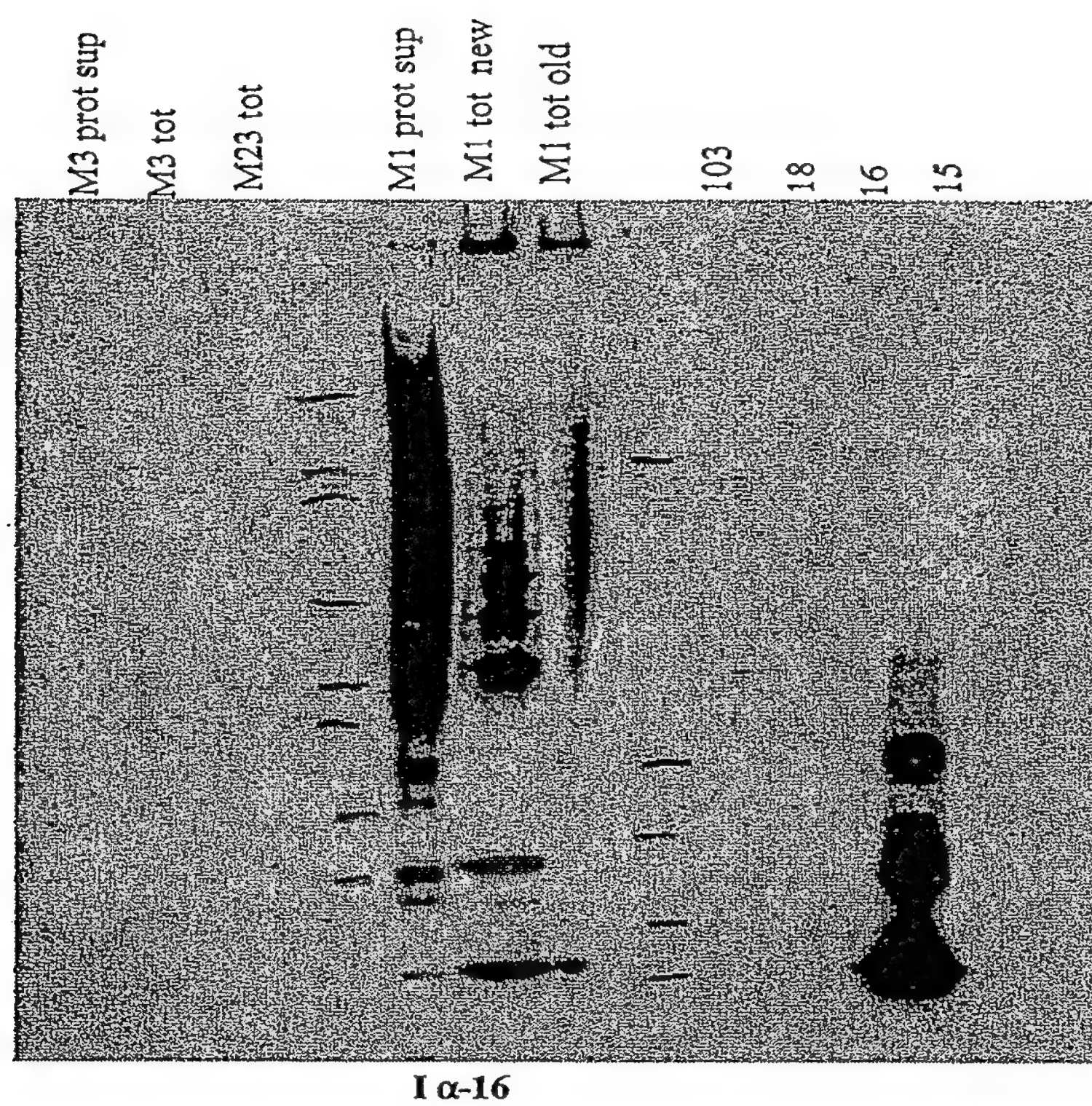
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 92

**LEGENDA:**

M1 tot: total extract (M1)

M1 prot. Sup.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

Figure 93

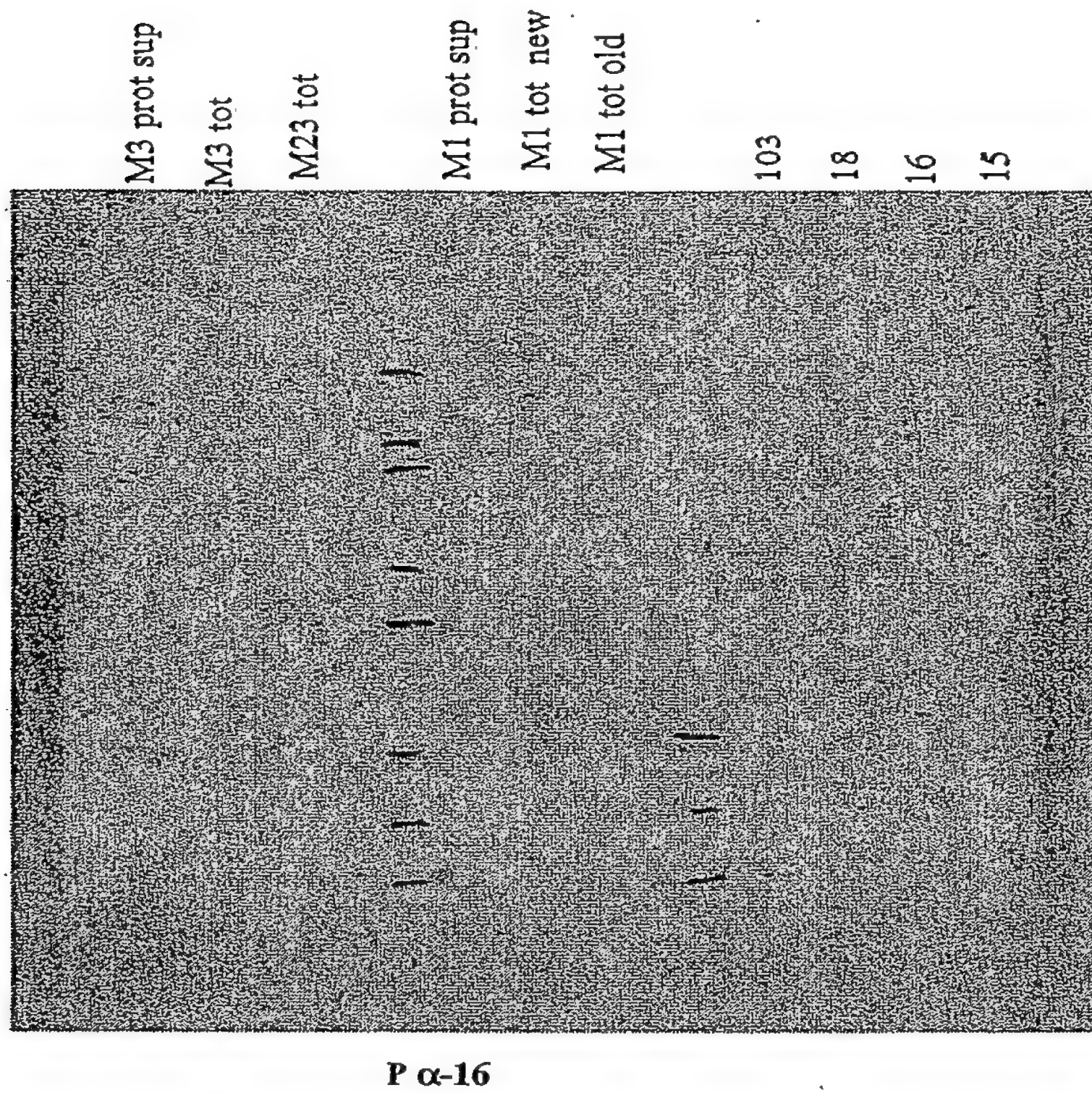
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 94

245/487

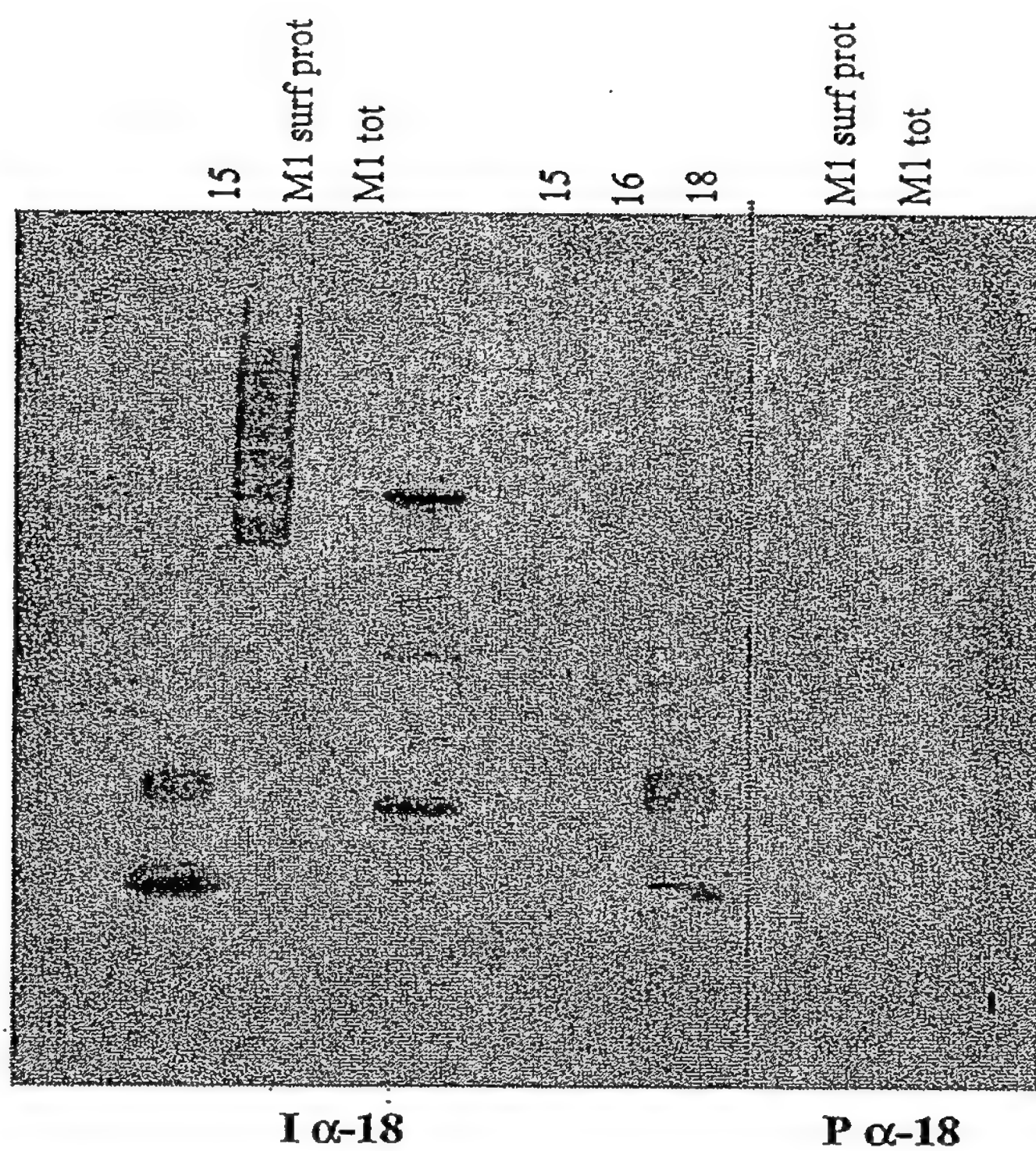
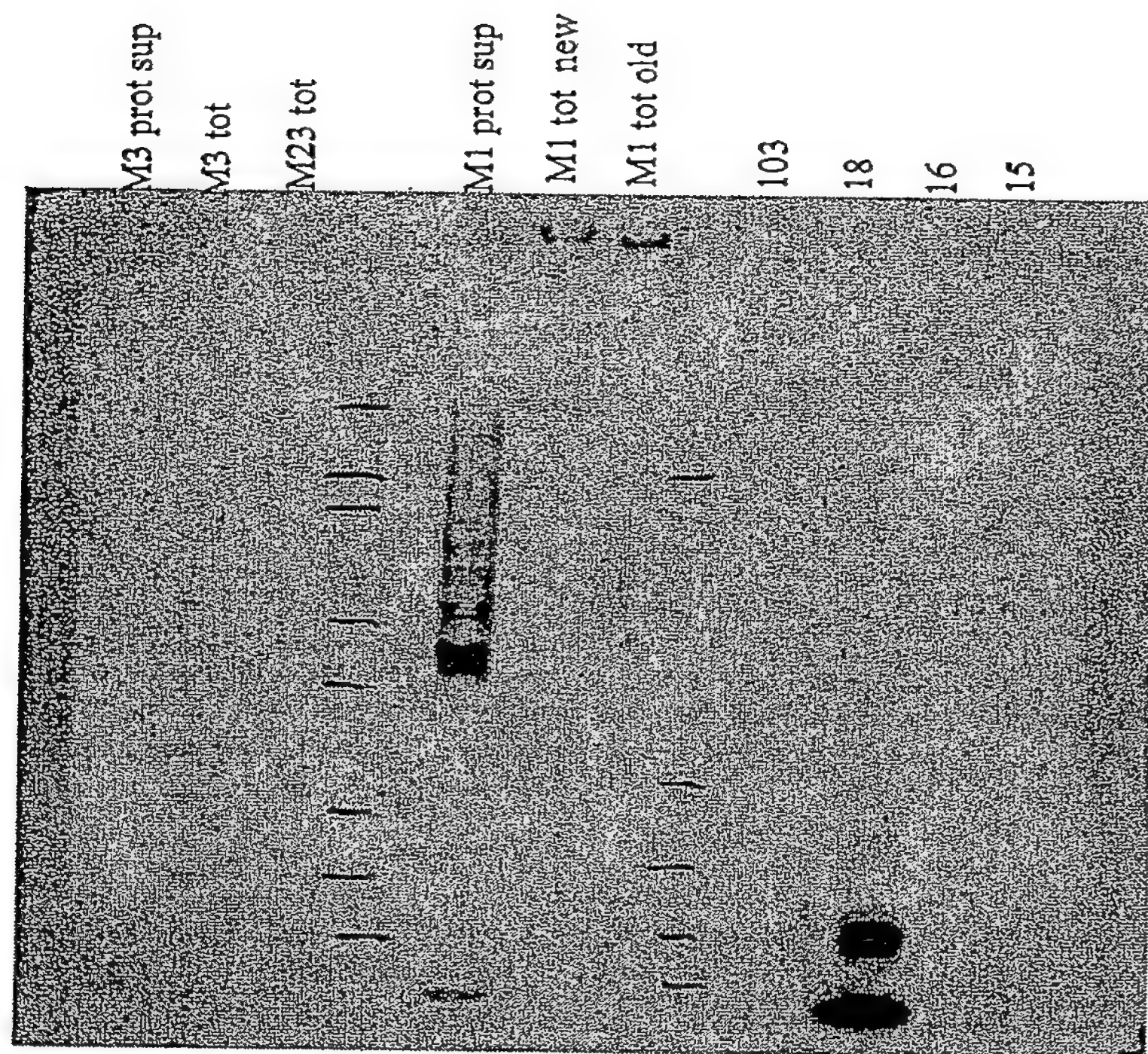
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 95

PCT/US05/27239 246/487

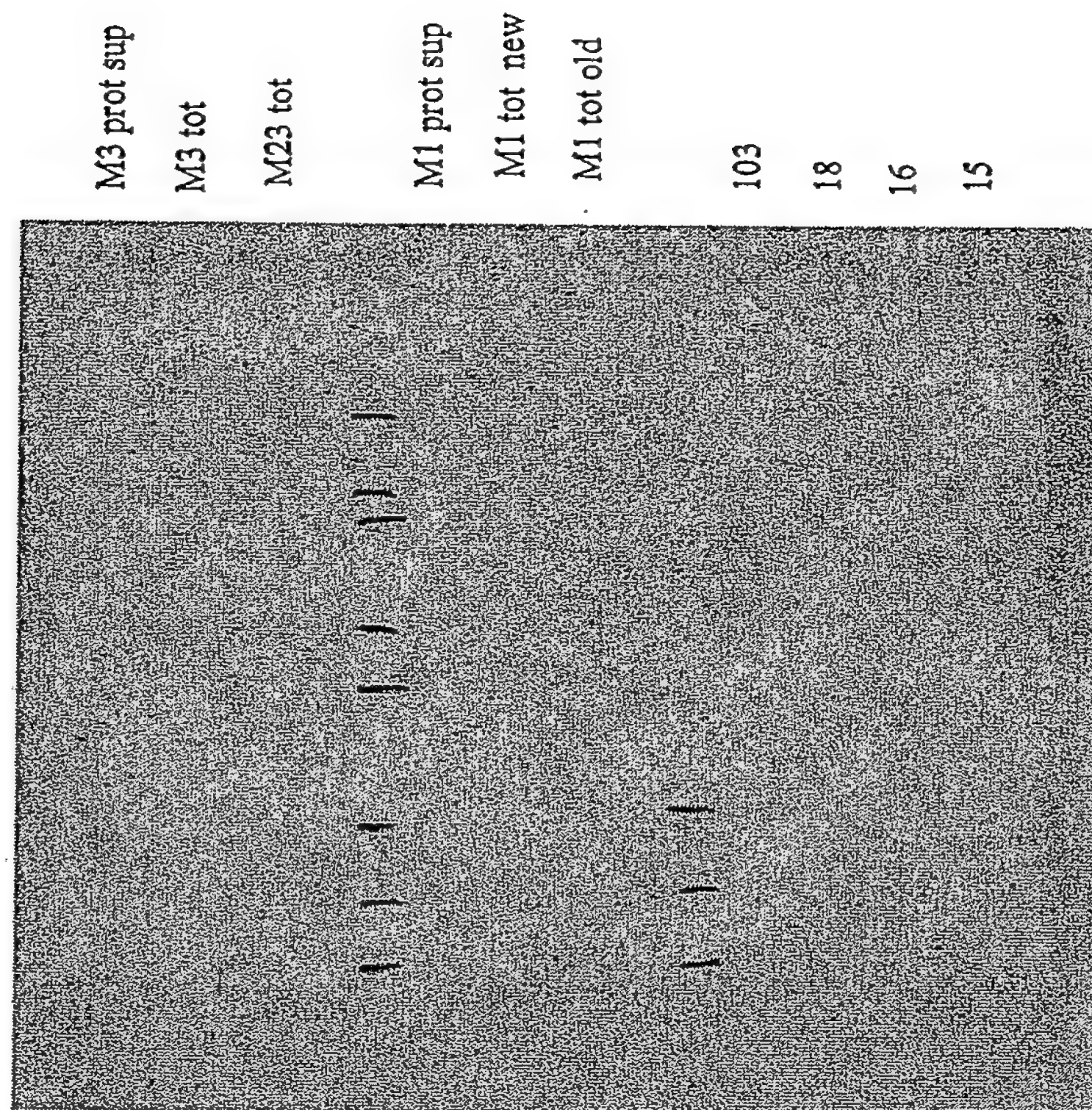


I α-18

LEGENDA:**M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 96

PC7/US05/27239 247/487



LEGENDA:

M1 tot: total extract (M1)

M1 prot. Sup.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

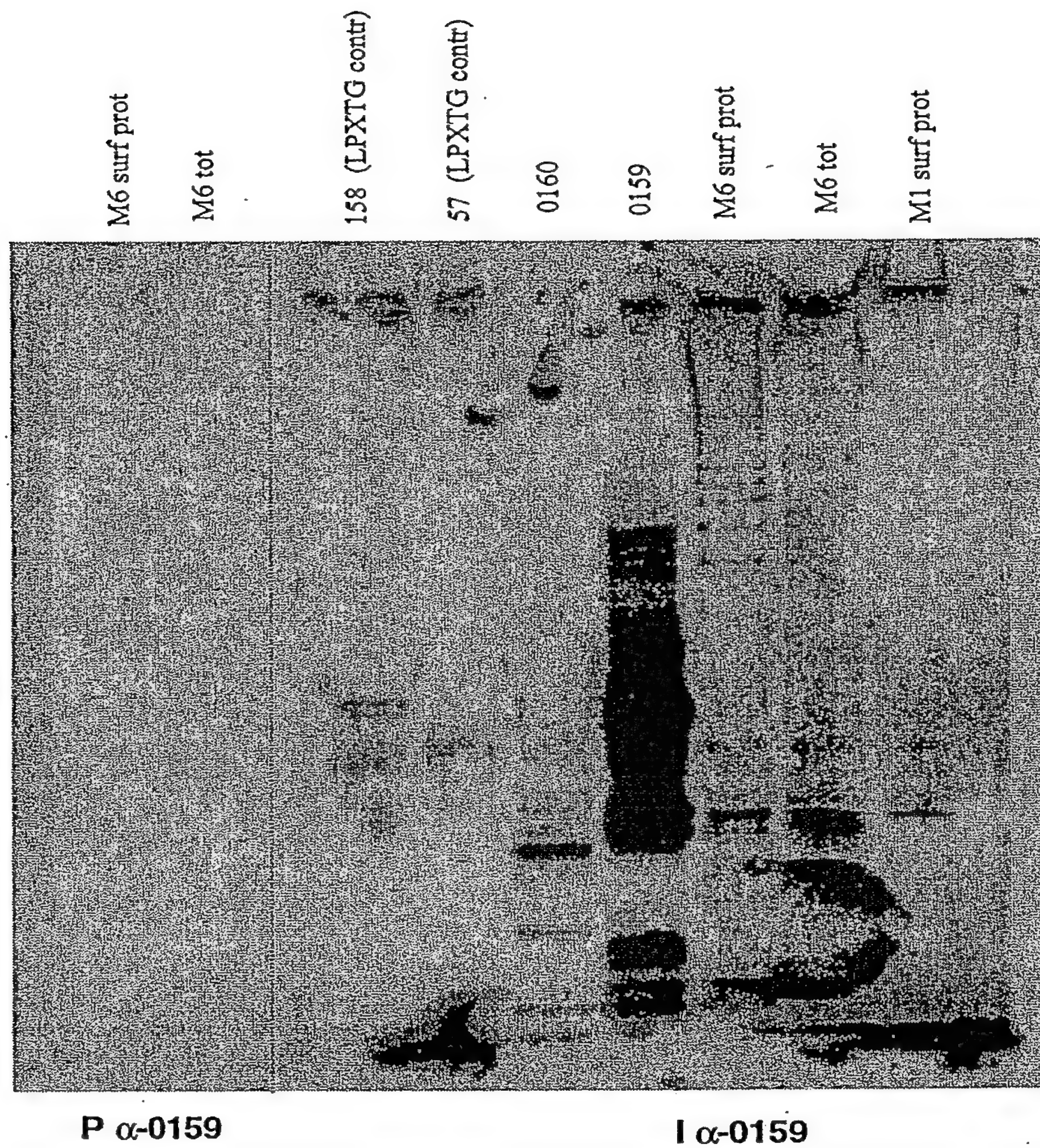
P α-#: pre-immune sera

P α-18

Figure 97

PCT/US05/27239 248/487

Figure 98

**LEGEND:**

M6 tot: total extract (M6)

M6 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

PCT/US05/27239 249/487

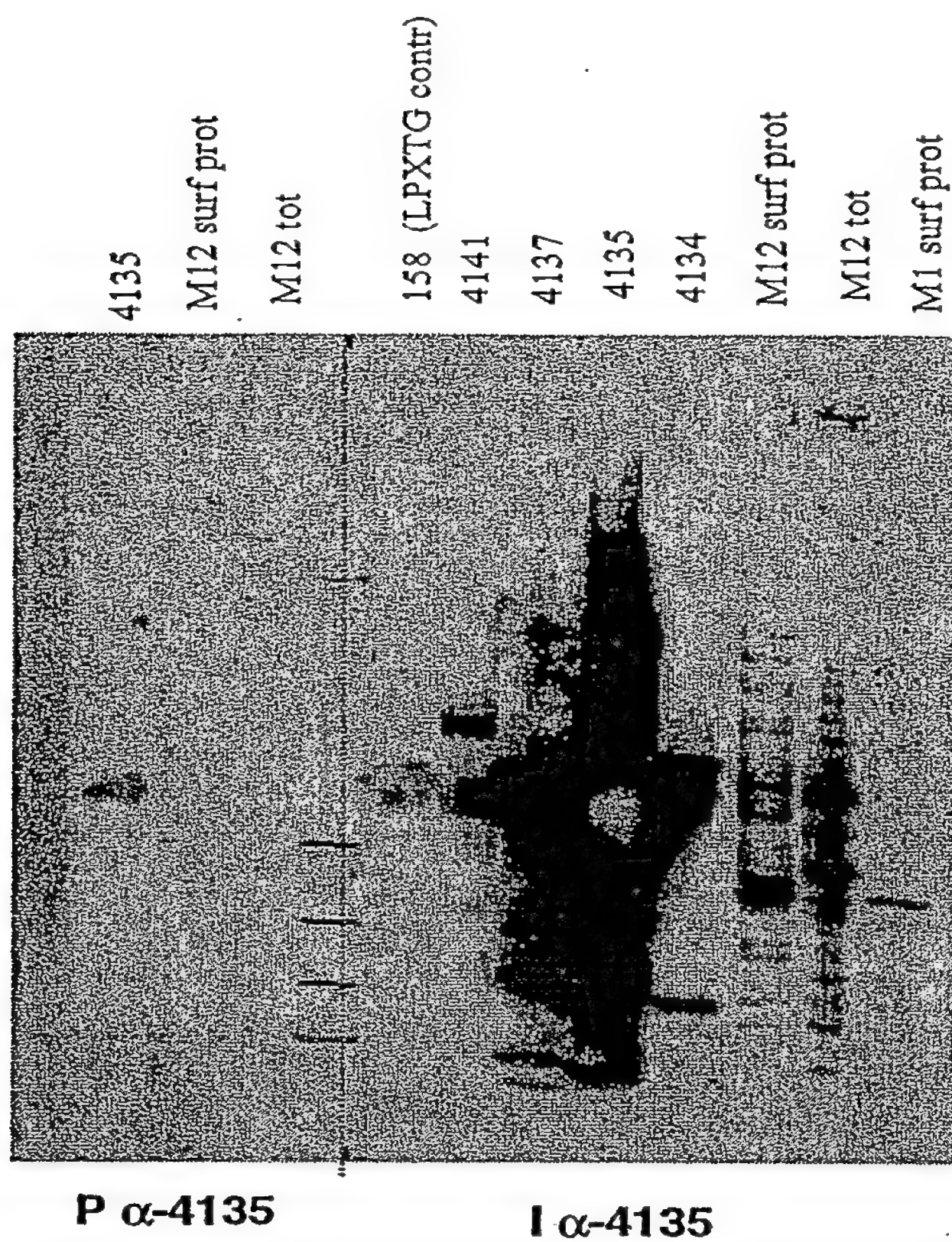
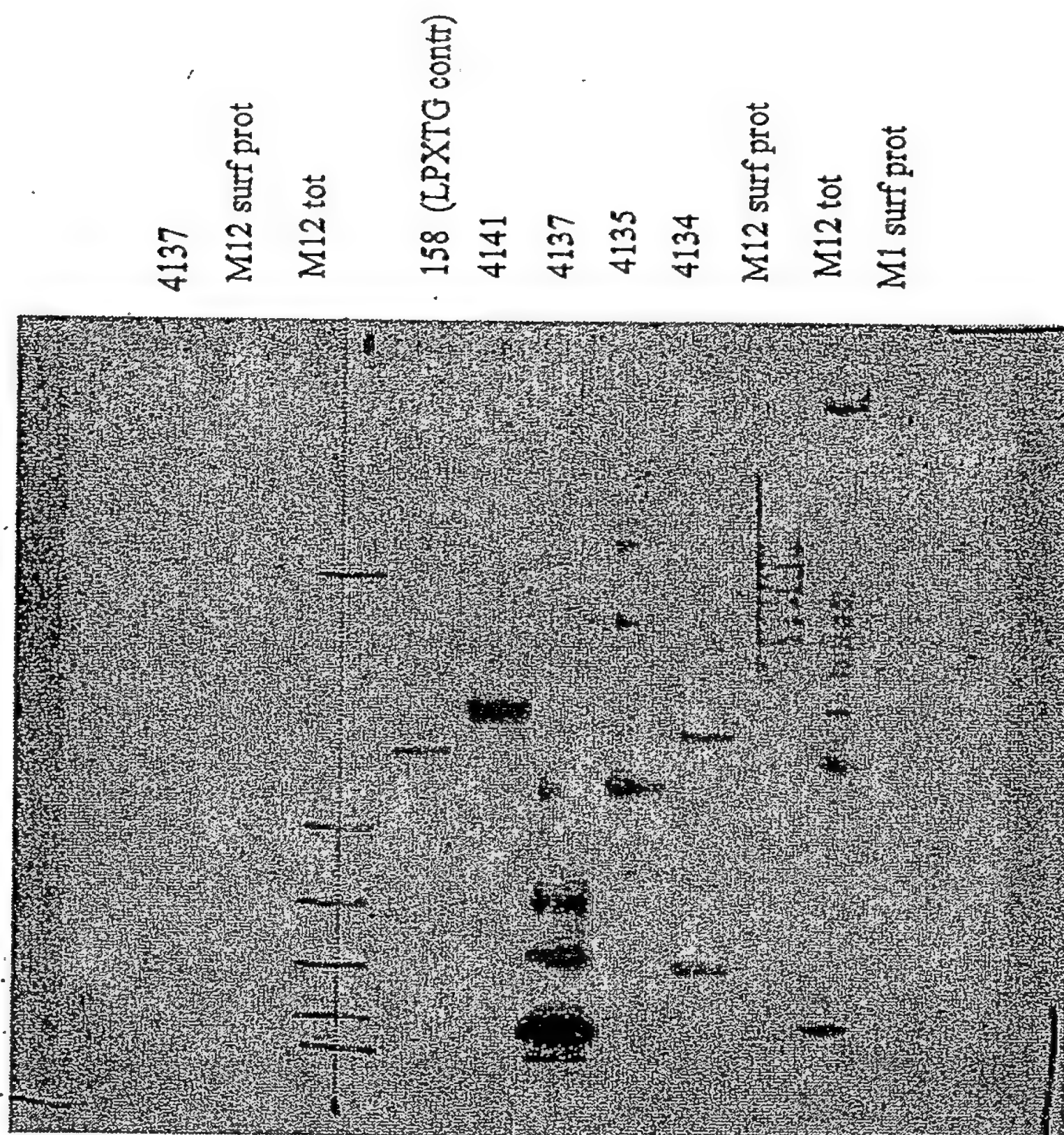
**LEGEND:****M12 tot:** total extract (M12)**M12 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 99

PCT/US05/27239 250/487

**LEGEND:**

M12 tot: total extract (M12)

M12 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α -#: immune sera against #P α -#: pre-immune seraP α -4137I α -4137

Figure 100

PCT/US05/27239 251/487

FIGURE 101

```
1  GACAAGCTTCCCTTATACGACCGCTTCTCTATATACGGACTCTTCCAAAGTTCCTGCTCTTTTA  60
   +-----+-----+-----+-----+-----+-----+
   CTGTTCGAAGGAATATGCTGGCGAAAGATATAGCCCTGAGAAGGTTCAAGAACGAGAAAAT

61  CCAGGGACTAGCCGTTCAAGGTGCAACGATGTGCGGTGTTGTAAATGGAACCAAGTCGT  120
   +-----+-----+-----+-----+-----+-----+
   GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAACAATTTACCTTGGTCAGCA

121 TCAGTTGTGACAGAAATTACCTTCTATCTTGGGATTCCTGTTATGTTTGGAGCTAGTGCC  180
   +-----+-----+-----+-----+-----+-----+
   AGTCAACACTGTCTTAAATGGAAGATAGAACCCCTAAGGGCAATACAAACCTCGATCACGG

181 TTAAAGATTTTCAAATTTGTGAAAGCCGGAGAACTCTTTGAGCTTTGGGCAATGTGTTTG  240
   +-----+-----+-----+-----+-----+-----+
   AATTCTTAAAGTTTAAACACTTTTCGGCCCTCTTTGAGAACTCGAAACCCGTTAACAAAAAC

241 CTC TTGGTCGGGATGGGAGTAGCTTTTGGCGGTCAGCATGGTGGCTATTCGCTTCTTGACC  300
   +-----+-----+-----+-----+-----+-----+
   GAGAACCAAGCGCTACCCCTCATCGAAACGCCAGTCGTACCAACCGATAAGCGAAGAACTGG

301 AGCTATGTGAAAAAACACGACTTCACCCCTTTTGGTAAATACCGTATCGTCTTGGTAGT  360
   +-----+-----+-----+-----+-----+-----+
   TCGATACACTTTTGTGTGCTGAAGTGGGAAAAAACCATTTATGGCATAGCACGAACCATCA

361 GTTTTGCTACTTTACAGTTTGTGTCGGTTTATTTGTATAAGAAAAACCTTGAAGGGTAAC  420
   +-----+-----+-----+-----+-----+-----+
   CAAAACGATGAATGTCAAACACAGGCAAAATAACATATTTCTTTTGGAACTTCCCCATTG

421 TCTTCAAGGTTTATACTCTTAGAAAAATCTCTTCAAACCGGTCAGCTTTATCTGCAACC  480
   +-----+-----+-----+-----+-----+-----+
   AGAAGTCCAAAATATGAGAAATCTTTTAGAGAAAGTTTGGCGCAGTCGAAATAGACGTTGG
```

Figure 101A

481 TC AAAACAGTGT TTTGAGCAGCCCTCGGCTAGCTTCTTAGTTTGCTCTTTGATTTTCATT 540
-----+-----+-----+-----+-----+-----+-----+
AGTTTGT CACAAAAC TCGTCGGACGCCGATCGAAGGATCAAACGAGAAACTAAAAAGTAA
541 GAGCTTTT AAAATCCAGTCAGGGTAATCCCAATAGCGGGACACCTCTTTCTTCTCGCTT 600
-----+-----+-----+-----+-----+-----+-----+
CTCGAAAATTTTAGGTCAGTCCCATTTAGGGGTTATCCGCCCTGTGGAGAAAGAAAGAGCGAA
601 AATTC TCATAGAGTTGCAGGGCTATTTGGCTTATCTGACTAGCATCTTGTTTGTGG 660
-----+-----+-----+-----+-----+-----+-----+
TTAAGAAGTATCTCAACGTCCCGATAAACCGAATAGACTGATCGTAGAACACAAAAAACC
661 CAAGACTTTTTCGTTTGGTAAGAGTTGAAAAGTCCTCGTAGCGGATTTTCAAAATGACAA 720
-----+-----+-----+-----+-----+-----+-----+
GTTCTGAAAAAGCAAAACCATCTCAACTTTTTCAGGAGCATCGCCTAAAGTTTACTGTT
721 TTTTTC CAGCTTTTCTTGTTGATGTAGATTGAGAGCGACTTTTCTGTAGAGAAGATCA 780
-----+-----+-----+-----+-----+-----+-----+
AAAAAGGTCGAAAAAGAACAACTACATCTA ACTCTCGCTGAAAAAGACTATCTTCTCAGT
840 GCTCTTTT TTGATATCTTCTCGGCACGGAGAACTCTCCCGTAGGTTTCTCTTCCCGCA 840
-----+-----+-----+-----+-----+-----+-----+
CGAGAAAAAACTATAGAGGAGCGCGTCCCTCTTAGAAGGGCATCCAAAAAGAGGAACGGCT
841 TTGATTTTACGGATGCGATTGGATTGTACTGGAGAGTTGTGAATGCCACGAGCCTTTCGAT 900
-----+-----+-----+-----+-----+-----+-----+
AACTAAATGCCCTACGCTAACCTAAACTGACCTCTCAACACTTACGGTGCTCGGAAAAGCTA
901 ACAGATCATAGCCTAGTCTACCAAAACGGTCTATTAGGGTTACCTCAGGAACITCAAGTA 960
-----+-----+-----+-----+-----+-----+-----+
TGTC TAGTATCGGATCAGATGGTTTGGCCAGATAATCCCAATGGAGTCCCTTGAAGTTCAT

PCT/US05/27239

253/487

Figure 101B

961 AATCAGCACAGTAAACGCGCCATTTGATGAAGACGTTCTACTGCTCTTTTTCCTACTC 1020
-----+-----+-----+-----+-----+-----+-----+
TTAGTCGTGGTCAATTTTGGGGTAAACTACTCTCTGCAAGATGACAGAAAAAGGATGAG

1021 CATGAAATTTGGAAATATCCATTTGTTGAGAAAAATCCTCAGCCTGTTTCAGGTAGAAATCA 1080
-----+-----+-----+-----+-----+-----+-----+
GTACTTTAAACCTTTATAGGTAACAAACTCTTTTAGGAGTCGGACAAAGTCCCATCTTAGT

1081 CTGTCAAACCATGTGGTTTGTGATAATCACTCGCCATTTTAGCTAAGAAATTTGTTGTAAG 1140
-----+-----+-----+-----+-----+-----+-----+
GACAGTTTGGTACACCAAAACCTATTAGTGAGCGGTAATAATCGATCTTTAAACAACATTC

1141 AAACGCCCTGGGGAAGCAGTTAGATGGAGTTCTTTCCAGATATCTTTTGAATGAGCGGAG 1200
-----+-----+-----+-----+-----+-----+-----+
TTTGGGACGCCCTTCGTCAATCTACCTCAAGAAAGGTCATATAGAAAAACTTACTCCGCTC

1201 CAATTTTGACCGCTGACTGTGATACCGAGTTTATTTTCTGTCAATCCAAATAGGCTTCGT 1260
-----+-----+-----+-----+-----+-----+-----+
GTTAAACTGGCGACTGAACCTATGGCTCAATAAAGACAGTGTAGGTTTATCCGAAGCA

1261 CAATGCTCATGGGTTCAATCAATCTGTATAGCGCTTAAAAATAGCTCGAATCCGGAGTC 1320
-----+-----+-----+-----+-----+-----+-----+
GTTACGAGTACCCAAAGTTAGTTTAGACATATCGCGAATTTTATCGAGCTTAGGCCCTCAG

1321 CCACAGACTTGTATTTCTCATAATTCCTGAGATAAAGACAGCCTGGGGACAACGTTTCAT 1380
-----+-----+-----+-----+-----+-----+-----+
GGTGTCTGAACATAAAGAGTATTAAAGGACTCTATTTCTGTCTCGGACCCCTGTTGCAAGTA

1381 AAGCTTCCCTTGGAACCTCATGGCAGAAATGGACACCAAAAGCTCTTGCCCTCATAACTACAGG 1440
-----+-----+-----+-----+-----+-----+-----+
TTCGAAGGAACCTTGAGTACCGCTCTTACCCTGTGGTTTTCGAGAACGGAGTATTGATGTCC

PCT/US05/27239

254/487

Figure 101C

1441 TAGAAACGACTCCCGTCCACCTGTTTGGCCGAGGTCGCTTCCAAATAATGACAGGTTTTC
-----+-----+-----+-----+-----+
1500 ATCTTTGCTGAGGGCAGGTGGACAAACGGCTCCAGCAAGGTATATTACTGTCCAAAAG
-----+-----+-----+-----+-----+
1501 CTCTGAGTTTAGGATTATCCCTGATTTCCACTGCAGCAAAAAGGCATCCATGTCATAT
-----+-----+-----+-----+-----+
1560 GAGACTCAAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTCCTGAGGTACAGTTATA
-----+-----+-----+-----+-----+
1561 GGATGATTTTCTTGACAAATCATTTAACAAGGAAAAATCAACATGCCCTAGCACCTTTT
-----+-----+-----+-----+-----+
1620 CCTACTAAAAAGAACTGTTTAGTAAATTGTTTCTTTTAGTGTACGGATCGTGGAAAA
-----+-----+-----+-----+-----+
1621 TATACTCTTCGAAAAATCTCTTCAACCACGTCAGCTTCCATCTGCAACCTCAAAACAGTA
-----+-----+-----+-----+-----+
1680 ATATGAGAAGCTTTTAGAGAAAGTTTGGTGCAGTCGAAGGTAGACGTTTGGAGTTTGTTCAT
-----+-----+-----+-----+-----+
1681 TTTTGAGCTGACTTCGTCAGTTCTATTTACAACCTCAAAGCAGTGCTTTTGAGCAGCCGCG
-----+-----+-----+-----+-----+
1740 AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTACGAAACTCGTCGGGACG
-----+-----+-----+-----+-----+
1741 GGCTAGTTTCCTAGTTTGCTTTTCGATTTCCATTTGAGTGTAACTGCTTATTTCTTTTAT
-----+-----+-----+-----+-----+
1800 CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACCTCACATTTGACGAATAAAGAAAAATA
-----+-----+-----+-----+-----+
1801 TATACCCCTTTTCTGAAAAAAGAAAAAGGACTTTATTTTTCAAAAATAATAATA
-----+-----+-----+-----+-----+
1860 ATATGGGAAAAAAGACTTTTCTTTTCTTTTTCCTGAAAAATAAAGTTTATATATATGT
-----+-----+-----+-----+-----+
1861 GTTTGAAAAATAATAGACTGTTTGTAGAAAAAGAGTGTAAAAATAGGAATTTTTCACCT
-----+-----+-----+-----+-----+
1920 CAAACTTTATTTTATATCTGACAAAAATCTTTTCTTTTTCACATTTTATCCTTAAAGTGA
-----+-----+-----+-----+-----+

Figure 101D

1921 TGTTGAAATCGGTTACTTTATGGTATATACTTGCTCATGAATGTAACAGATGACTGTTACT + 1980
ACAACTTTAGCCCAATGAAATACCATATGAACAGAGTACTTACATTTGCTACTGACAATGA
1981 AGAAAAAGAGGACATTAATATGGTTGTTAAGACAGTTGTTGAAGCACAAAGATATTTTG + 2040
TCTTTTTCCTGTAATTATATACCAACAATTCTGTCAACAACCTTCGTTCTATATAAAAC
M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate
acetyltransferase (pfl). It is out of the
pilus locus
2041 ACAAGCTTGGGAAGGCTTCAAGGCGTAGATTGGAAGAAAAAGCAAGTGATCACGCT + 2100
TGTTTCGAACCCCTTCGGAAGTTTCCGCATCTAACCTTCTTTTTCGTTCCACATAGTGGGA
K A W E G F K G V D W K E K A S V S R F -
2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAGCTTCTTGCAGGACCAACAG + 2160
AACATGTTCCGATTGATGTTGGGAATACTACCTCTGCTTTTCGAAGGAACGTCCTGGTTGTC
V Q A N Y T P Y D G D E S F L A G P T E -
2161 AGCGTTCACATCAAGAAAATTGTAGAAGAACTAAAGCACACTACGAAGAACTC + 2220
TCGCAAGTGAAGTGTAGTTCTTTTAAACATCTTCTTTTGATTTTCGTTGATGCTTCTTTGAG
R S L H I K K I V E E T K A H Y E E T R -
2221 GTTTCCTCAATGGACACTCGTCCACATCTATCGCTGATATCCCTGCTGGATTATCGACA + 2280
CAAAGGTTACCTGTGACAGGTTGTAGATAGCGACTATAGGACGACCTAAATAGCTGT
F P M D T R P T S I A D I P A G F I D K -
2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTCAAAATGAACCTCATGC + 2340
TTCCTTTTACTTCAATAGAAACCATAGGTTTACTACTTGAGAAAGTTTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -
2341 CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAGAAATGGATACGAACAGACC
GTTTCCACCATAGGCATACCGACTTTGATGAAATTTCTTTTACCTATGCTTGGTCTGG 2400
C K G G I R M A E T T L K E N G Y E P D P -
2401 CAGCTGTTACGAAATCTTCACTAAATATGTAAACAACAGTTAACGACGGTATTTCCCGTG
GTCGACAAGTGCCTTTAGAGTGATTTATACATTGTGTCAATTGCTGCCATAAAAGGCAC 2460
C A V H E I F T K Y V T T V N D G I F R A -
2461 CCTACACTTCAATATTCGTCGGCTCGTCAATGCACACACTGTAACTGGTCTTCCAGATG
GGATGTGAAGTTTATAAGCAGCGGAGCAGTACGTGTGTGACATTGACCAGAAGGTCTAC 2520
C Y T S N I R R A R H A H T V T G L P D A -
2521 CATACTCAGCGGACGTATCATCGGTGTTTACGCACGTCCTTGTCTTTACGGTGCAGACT
GTATGAGTGGCCCTGCATAGTAGCCACAAATGCGTGCAGAACGAGAAATGCCACGCTCTGA 2580
C Y S R G R I I G V Y A R L A L Y G A D Y -
2581 ACTTGATGCAAGAAAAGTAAACGACTGGAAATGCAATCAAGAAATCGATGAAGAAACAA
TGAACACTAGTTCTTTTTCATTGTGCTGACCTTACGTTAGTTTCTTTTAGCTACTTCTTTGTT 2640
C L M Q E K V N D W N A I K E I D E E T I -
2641 TCCGTCTTCGTGAAGAAGTAAACCTTCAATACCAAGCATTTGCAACAAGTTGTTCCGCTGG
AGCAGAAAGCACTTCTTCAATTGGAAGTTATGGTTCTGTAACGTTGTTCAACAAGCGGACC 2700
C R L R E E V N L Q Y Q A L Q Q V V R L G -
2701 GTGACCTTTACGGGGTTGATGTTGCGAAACACGCGATGAACGTGAAGAAAGCAATCCAAT
2760

PCT/US05/27239 257/487

Figure 101F

```
CACTGGAAATGCCCAACTACAAGGTTTGGTCGCTACTTGCACCTTCTTCGTTAGGTTA
      D L Y G V D V R K P A M N V K E A I Q W -
2761  GGGTTAACATTGCTTTCATGGCTGCTGCGGTGATTAACGGTGCTGCTACATCTCTAG
      +-----+-----+-----+-----+-----+
      CCCAATTGTAAAGTAACCGACAGACGGCACACTAATTGCCACGAGATGTAGAGATC
      +-----+-----+-----+-----+-----+
      V N I A F M A V C R V I N G A A T S L G -
2821  GTCGTGTACCAATCGTATTGGACATCTTTGCAGAACGTCGCTGCTCGTGGTACATTTA
      +-----+-----+-----+-----+-----+
      CAGCACATGGTTAGCATAAACCTGTAGAAAACGTCCTTGCACTGGAACGAGCACCATGTAAAT
      +-----+-----+-----+-----+-----+
      R V P I V L D I F A E R D L A R G T F T -
2881  CTGAATCAGAAAATCCAAGAAATTCGTTGATGATTTTCGTTATGAAAACCTTCGTACAGTTAAAT
      +-----+-----+-----+-----+-----+
      GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATGTCAATTTA
      +-----+-----+-----+-----+-----+
      E S E I Q E F V D D F V M K L R T V K F -
2941  TTGCTCGTACCAAGCTTATGACCAATGTACTCAGGTGACCAACCTTTATCACAACCTT
      +-----+-----+-----+-----+-----+
      AACGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACCTGGGTTGGAAATAGTGTGAA
      +-----+-----+-----+-----+-----+
      A R T K A Y D Q L Y S G D P T F I T T S -
3001  CTATGGCTGTTATGGGTAACGACGGTCGTCACCGTGTACTAAGATGGACTACCGTTTCT
      +-----+-----+-----+-----+-----+
      GATACCGACCATACCCATTGCTGCCAGCAGTGGCACAAATGATTCTACCTGATGGCAAAGA
      +-----+-----+-----+-----+-----+
      M A G M G N D G R H R V T K M D Y R F L -
3061  TGAACACTCTTGACAACATGGGTAACCTCACCAGAACCAAACTTGACAGTTCCTTTGGACTG
      +-----+-----+-----+-----+-----+
      ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAACCTGTCAAGAAACCTGAC
      +-----+-----+-----+-----+-----+
      N T L D N I G N S P E P N L T V L W T D -
```

PCT/US05/27239 258/487

Figure 101G

3121 ~ACAAATTGCCATACAACTTCGGTCGCTACTGTATGCACATGAGCCACAAACACTCTTCTA 3180
-----+-----+-----+-----+-----+-----+-----+
TGTTTAACGGTATGTTGAAGGAGCGGATGACATACGTTACTCGGTGTTTGTGAGAAAGAT
K L P Y N F R R Y C M H M S H K H S S I -
3181 TCCAATACGAAGGTGTAACAACAATGGCTAAAGACGGATATGGTGAAATGAGCTGTATCT 3240
-----+-----+-----+-----+-----+-----+-----+
AGGTTATGCTTCCACATGTTGTTTACCGATTCTGCTATATACCACCTTTACTCGACATAGA
Q Y E G V T T M A K D G Y G E M S C I S -
3241 CATGCTGTGTCTCCACTTGATCCAGAAATGAAGAACAACGCCACACATCCAGTACT 3300
-----+-----+-----+-----+-----+-----+-----+
GTACGACACACAGAGGTGAACCTAGGTCCTTTACTTCTTGTGCGGTGTTGTAGGTCATGA
C C V S P L D P E N E E Q R H N I Q Y F -
3301 TCGGTGCTCGTGTAACGTTCTTAAAGCCCTTCTTACTGGTTTGAATGGTGGTTACGACG 3360
-----+-----+-----+-----+-----+-----+-----+
AGCCACGAGCACATTTGCAAGAATTCGGGAAGAATGACCAAACTTACCACCAATGCTGC
G A R V N V L K A L L T G L N G G Y D D -
3361 ATGTTCAAAAGACTACAAAGTATTTGATATCGAACCAATCCGGTGACGAAGTTCTTGAAT 3420
-----+-----+-----+-----+-----+-----+-----+
TACAAAGTGTTCCTGATGTTTCATAAATACTATAGCTTGGTTAGGCACCTGCTTCAAGAACTTA
V H K D Y K V F D I E P I R D E V L E F -
3421 TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTGACTGGTTGACTGACACTTACGTTAG 3480
-----+-----+-----+-----+-----+-----+-----+
AACTTAGTCAATTCGCTTGAAACTTTTATAGAGAAGTACCAACTGACTGTGATGATC
E S V K A N F E K S L D W L T D T Y V D -
3481 ATGCCCTTGAACATCATCCACTACATGACTGATAGGTACAACCTACGAAGCTGTCAATGG 3540
-----+-----+-----+-----+-----+-----+-----+
TACGGAACTTGTAGTAGGTGATGTACTGACTATCCATGTTGATGCTTCGACAAAGTTTACC

PCT/US05/27239 259/487

Figure 101H

C A L N I I H Y M T D R Y N Y E A V Q M A -
3541 CCTTCTTGCCAACTAAACAAGTGCCCAACATGGGATTCGGTATCTGTGGATTGCTAACA + 3600
GGAAGAACGGTTGATTGTTGTCACGGTTGTACCCCTAAGCCATAGACACCTAAACGATTGT
C F L P T K Q R A N M G F G I C G F A N T -
3601 CTGTTGATACATTGTCAGCTATCAAAATACGCTACAGTTAAACCAATCCGTGACGAAGATG + 3660
GACAACTATGTAACAGTCGATAGTTTATGCGATGTCAATTTGGTTAGGCACCTGCTTCTAC
C V D T L S A I K Y A T V K P I R D E D G -
3661 GCTACATCTACGATTACGAAACAATCGGTGACTACCCACGCTGGGTGAAGATGACCCAC + 3720
CGATGTAGATGCTAATGCTTTGTTAGCCACTGATGGGTGGACCCCACTTCTACTGGGTG
C Y I Y D Y E T I G D Y P R W G E D D P R -
3721 GTTCAAACGAATTTGGCAGAAATGGTTGATCGAAGCTTACACAACCTCGTCTACGTAGCCACA + 3780
CAAGTTTGCTTAACCGTCTTACCACTAGCTTCCGAATGTGTGAGCAGATGCATCGGTGT
C S N E L A E W L I E A Y T T R L R S H K -
3781 AACTATACAAAGACGCAAGCTACAGTATCATTTTTGACAATCACAATCAACGTTGCTT + 3840
TTGATATGTTTCTGCGTCTTCGATGTCATAGTGAAACCTGTTAGTGTAGATTGCAACGAA
C L Y K D A E A T V S L L T I T S N V A Y -
3841 ACTCTAAACAACCTGTAACCTCACAGTTCACAAAGGTGTATACCTCAACGAAGATGGTT + 3900
TGAGATTGTTTGACCATTTGAGTGGTCAAGTGTTCACACATATGGAGTTGCTTCTACCAA
C S K Q T G N S P V H K G V Y L N E D G S -
3901 CTGTGAACCTTGTCTAAACCTTGAATTTCTCTCACCAGGTGCTAACCCATCTAACAAAGCTA + 3960

C		V	N	L	S	K	L	E	F	F	S	P	G	A	N	P	S	N	K	A	K	-
		GACACTTGAACAGATTTGAACCTTAAGAAGAGTGGTCCACGATTGGGTAGATTGTTTCGAT																				
		AAGGTGGTGGTTGCCAAAACCTTGAACTCACCCTTCTAGCCTTGACTTTAGTTATGCAGCTG																				
3961		TTCCACCAACCAACGTTTTTGAACCTTGAGTGAAGATCGGAAGTGAATCAATACGTCGAC																				
		G G W L Q N L N S L S S L D F S Y A A D -																				
		ACGGTATCTCATTGACTACACAAGTATCACCTCGCGCTCTTGGTAAGACTCGTGATGAAC																				
4021		TGCCATAGAGTAACTGATGTGTTTCATAGTGGAGCGCGAGAACCATTTCTGAGCACTACTTG																				
		G I S L T T Q V S P R A L G K T R D E Q -																				
		AAGTTGATAACTTGGTAACAATCCTTGATGGTTACTTTCGAAAACGGTGGACAAACACGTTA																				
4081		TTCAACTATTGAACCATTTGTTAGGAACCTACCAATGAAGCTTTTGGCCACCCTGTTGTGCAAT																				
		V D N L V T I L D G Y F E N G G Q H V N -																				
		ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTCAGGCGAAGACGTTA																				
4141		TGAACCTGCAATACCTGAACCTTGCTACAAATGCTTTTTTTAGTACAGTCCGCTTCTGCAAT																				
		L N V M D L N D V Y E K I M S G E D V I -																				
		TCGTACGTATCTCTGGATACTGTGTAAACACTAAATACCTCACCAGAACAAAAAAGCTG																				
4201		AGCATGCATAGAGACCTATGACACATTGTTGATTATGGAGTGAGGTCTTGTGTTTTTTTGAC																				
		V R I S G Y C V N T K Y L T P E Q K T E -																				
		AATTGACACAACGTTCTTCCACGAAGTTCTTTCAATGGATGACGCCCTTGGATGCATTGA																				
4261		TTAACTGTGTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAAC																				
		L T Q R V F H E V L S M D D A L D A L S -																				

PCT/US2005/027239

261/487

Figure 101J

4321 GCTAATCAAGTCTTGAATAATAAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA + 4380
-----+-----+-----+-----+-----+-----+
CGATTAGTCAAGAACTTATTTATTTCCCGAGAAACAGTTGACATCACCCAACTTCTTTT
c *
4381 AGCTAAGCTCGAGAAAGGACAAATTTGTCTTTCTTTTGTGATGTTTCAGAGCGATGAAA + 4440
-----+-----+-----+-----+-----+-----+
TCGATTCGAGCTCTTTCCTGTTTAAACAGGAAAGAAAACTACAAGTCTCGCTACTTTT
e * A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase
4441 ATCCGTTTTTTGAAGTTTTCAAGTTCGGAACCAAGGCAATGCGCTGTGATGCTTTTG + 4500
-----+-----+-----+-----+-----+-----+
TAGGCAAAAAAACTTCAAAAGTTTCAAGGCTTTTGGTTTCCGTAACGGGAACACAGAAAC
e I R K K F N E F N R F G F A N R K I D K -
4501 ATGAGTTTGTAGTGGCTCAAGTTTAGCGTTAGAATAAGGCAATTCAATGGCGTTAGTG + 4560
-----+-----+-----+-----+-----+-----+
TACTCAACAATCACCGAGTTCAAAATCGCAATCTTATTCCGTTAAGTTACCGCAATCAC
e I L K N T A E L K A N S Y P L E I A N T -
4561 ATGTAGTTTTTATAGCAATAAATGTGCTCAAGTGGTTTAAAGGTGCGGTGAGATGA + 4620
-----+-----+-----+-----+-----+-----+
TACATCAAAAAATATCGTTTATTATACACGAGTTTCACCAAAATTTCCACGCCAATCTACT
e I Y N K Y C I F T S L T T K F T R N L H -
4621 GGTAACGTGTCTTGAATTAAGCCCCAAACTGGTCAGTATTTCTCTGTAGATGAAT + 4680
-----+-----+-----+-----+-----+-----+
CCATTGCACAGAACTTAATTCGGGGTTTGGACCATATAAGAGAGAACATCTACTTTA
e P L T D Q I L G W F Q D T N K E Q L H F -
4681 AGGAGTAGTTGATACAGGTCATAGTAATCTTTAAGTTTCAGGTACTAGAGTAAAGATTTTC + 4740
-----+-----+-----+-----+-----+-----+
TCCTCATCAACTATGTCAGTATCATTAGAAATTCAGTCCATGATCTCATTTCTTAAAG

262/487

Figure 101K

```

e L L L Q Y L D Y Y D K L E P V L T F I K -
4741 TT CAGACACTCC TAGAGTTAAGGTCCTCTCTGAAAGTTCTAGCATAGAAAAGGCTTAAGA
+-----+-----+-----+-----+-----+-----+
AAGTCGTGAGGATCCCTCAATCCAGAGAGACTTTC AAGATCGTATCTTTCGAAATTCT
4800
e K L C E R P T L T E R F T R A Y F P K L -
4801 GAGAGTTCCGACTATCTTTTAGGATAAAATTTCCAGTAATATTTAAGAGCTCTGTATTCC
+-----+-----+-----+-----+-----+-----+
CTCTCAAAGGCTGATAGAAAATCCTATTATTTAAAGGTCATTATAAATCTCGAGACATAAGG
4860
e S L K R S D K L I F K W Y Y K L A R Y E -
4861 AGAGATTTATCATCAAATTGCTTCATGATGTTGATTCTAGTCTGATTAAGAGCCCTGCTC
+-----+-----+-----+-----+-----+-----+
TCTCTAAATAGTAGTTTAAACGAAGTACTACAACTAAGATCAGACTAATCTCTCGGACGAG
4920
e L S K D D F Q K M I N I R T Q N L A R S -
4921 ATGTGTTGGACAATGTGGAACGATCGAGAACAAATTTTAGCATGGGAAATAATTCTTA
+-----+-----+-----+-----+-----+-----+
TACACAACCTGTTACACCTTTGCTAGCTCTTGTTAAATCGTAACCTTTATTATAAGAAT
4980
e M H Q V I H F R D L V I K A N P F L K K -
4981 ATGAGAGGGATATAACTTCCAGACATATCAACAGTGACGACTTTAACTTTTCTTAGCT
+-----+-----+-----+-----+-----+-----+
TACTCTCCCTATATTGAAGGCTGTGTATAGTTGTCTGCTGAAATTTGAAAAAAGATCGA
5040
e I L P I Y S G S M D V T V V K V K K R A -
5041 TCCTTCGAGTACTGAAGAAATGATTTCGGATGGTTGTTGACGCTCTGTTATCAAGAATG
+-----+-----+-----+-----+-----+-----+
AGAAAGCTCATGAACCTCTTTACTAAAGCCTACCAACAACCTGCAGACAAATAGTCTTAC
5100
e E K S Y K F F H N R I T T Q R R N D L I -
5101 GTCATGATTTTCTTAGTGTGAAATCCTGAGCAATGAAGCCAAATTTCCCTTCTGCTAG
+-----+-----+-----+-----+-----+-----+

```

Figure 101L

```
e      CAGTACTAAAGAATCACAACCTTTAGGACTCGTTACTTTCGGTTAAAGGGAAGACCATC
      T M I K K T N F D Q A I F A L K G K Q Y -
5161  GAGAAATTCATCCAGGAGAGGATTTAGGCAAGTGGTGAATCCTCTTGGAATGAAT
      -----+-----+-----+-----+-----+
      CTCTTAAGTAGGTCCTCTCTCTAAAGTCCGTTTACACATAGGAGAACCTTTACTTTA
      5220
e      S F E D W S L I E P L T T Y D E Q F H F -
      TGCTTGAGCTTACGATAGACGGTAGAGGTAGAGATGGCTAATTTAGAAGCGATA
      -----+-----+-----+-----+-----+
      ACGAACTCGAATGCTATCTGCCATCTCCATCTCCATCTCTACCGATTAATCTTCGCTAT
      5280
e      Q K L K R Y V T S T S T S I A L K S A I -
      TGTGTAAGAGCCCTCTCTGTGTAGTAGGAGTTGGGCAATTTTCTGTCTCACCATTTCGGAG
      -----+-----+-----+-----+-----+
      ACACATCTCGGAGAGACAACTCATCTCAACCCGTTAAAGACAGAGTGGTAAAGGCTC
      5340
e      H T L A E R N L L L Q A I K Q R V M E S -
      ATTTGGCAATTTTCTGAACGAGAGTTGTTTCAGCTACAGTGACTTTCGACAGGACTTG
      -----+-----+-----+-----+-----+
      TAAACCGTTAAAGAGACTTGCTCTCAACAAGTCGATGTCACGAAAGGCTGCCCTGAAC
      5400
e      I Q C N K Q V L T F E A V T V K R C S K -
      CATTGAAATCGTCTCTTTTCAAAATGAATGAGGCTAGGGAACCAACCAATCTCGATAAAA
      -----+-----+-----+-----+-----+
      GTAACTTTAGCAGAGAAAAGTTTACTTACTCCGATCCCTTTGGTGGTTAGAGCTATTTT
      5460
e      C Q F R R K K L H I L S P F G G I E I F -
      GGGATTTTAGAAGGCTTTTGGAGTCGTATTTGATTTGTTTTCCTTTTACAGTGTTTACAT
      -----+-----+-----+-----+-----+
      CCTAAAATCTTCCGAAAACCTTCAGCATAAACTAAACAAAAGGAATGTCACAAATGTA
      5520
e      P I K S P K Q F D Y K I Q K G K C H K C -
```

5521 TTAGGTGGGTGATAATCAAGTGTAGCGAAGACTTCGATATATGGGTATCGTCTGAATGGCT
5580
AATCCACCACCTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCACGACTTACCGA
K P P H Y D L T A F V E I H T D H Q I A -
5581 TTATTTAAGGTGATGTTTTTTTGTCTTTTATTCCGATGAGTAATGTGGTATGATGATGTGT
5640
AATAAATTCCACTACAAAACAGAAAATAAAGGCTACTCATACACCATACTAACTACACACA
K N L T I N K D K I G I L L T T H N I H -
5641 TCCATAAGATACTTTCTAATGAGTTGTTTAGGGCGCTTTTTCATTATAAGTCTTATGGGACT
5700
AGGTATTCTATGAAAGATTACTCAACAAATCCGGGAAAAGTAATATTCAGAAATACCCCTGA
E M
5701 TTTTGTGATACTCAAAAAGCCCTATAAATCTCCACAGTGGGATTTACCCACTACAGAAATTA
5760
AAAACTATGAGTTTTTTCGGGATATTAGAGGTGTCAACCCTAAATGGGTGATGCTCTTAAT
TAGAGCCAGAAAACACTTTTGTTCAC TAGCAGAACTAGAGAGCAGAAAGTGTTTTCT
5761
ATCTCGGTCTTTTTTGTGAAAACAAGTGATCGTCTTTTGATCTCTCGTCTTCACAAAAGA
5820
GTTCAGATTTACCCAAAAC TGGGAAATATG GGGATAAGAAATAGAGATGGCTTAGGAAGCC
5880
CAAGTCTAAATGGGTTTTGACCCCTTATACCCCTATTCTTATCTCTACCGAATCCCTTCGG
CCTTTTTTGTGTAGACAGTACGATGAACCTTATAACAATAAGTGAGCCCTTTTTAGCAATC
5881
GGAAAAACACACATCTGTCTATGCTACTTGAATATTGTTTATCATCTCGGAAAAAATCGTTAG
* L L Y H A K K A I -orf2
ATTGCGACCCGTTTGTCAAAAAGCCCTCTTTTTCGGATATCTACAATTGTCTGATAGATGAGA

PCT/US2005/027239

265/487

Figure 101N

```
5941 -----+-----+-----+-----+-----+-----+-----+ 6000
      TAACGCTGGGCAACAGTTTTCGGAGAAAAGCCCTATAGATGTTAACAGACTATCTACTCT
e      M A V R K D F A E K R I D V I T Q Y I L -
      CGCTGTTGGCTAACATGCAAAATCTAAGGCAATCGTCAAAAAGTGATGTTCCCTTTGGGA
6001 -----+-----+-----+-----+-----+-----+-----+ 6060
      GCGACAACCGATTGTACGTTTAGATTCCGTTAGCAGTTTTCACCTACAAAGGGAACCCCT
e      R Q Q S V H L D L A I T L F H H K G K P -
      TACTGCTTTTAAACGTAAGGAGGATTTCTTTCGTTGTAATAATAATCAATGGCTCTGTC
6061 -----+-----+-----+-----+-----+-----+-----+ 6120
      ATGACGAAAAAATTGCATTCCGTCCTCCATAAGAAAGCAACATTATTATTAGTTACCGAGACAG
e      Y Q K K V Y P L Y E K T T I I I L P E T -
      AAATGCTCCTCTGAAGGAGGAGGACTAATTAGAAATATTGTATCCTGTAAACAGAGGCAACT
6121 -----+-----+-----+-----+-----+-----+-----+ 6180
      TTTACGAGGAGACTTCTCCTCCTGATTAATCTTATAACATAGGACATTGTCTCCGTTGA
e      L H E E S P P P S I L I N Y G T V S A V -
      TTGTCAGTAAATTCGGTAAATAATGGACTTTTATTAAAGTTTACATCTCCTTGATTATTT
6181 -----+-----+-----+-----+-----+-----+-----+ 6240
      AACAGTCATTTTAAAGGCATTTTATTATACCTGAAATAATCAAAATGTAGACGAACATAATAA
e      K D T F N R L I I S K I L N V D A Q N N -
      AAAATGATAAAAATCGGGATAGCAGGTAGTGAGGAAAAGATGGTTTCTGTCAAGTAGAGT
6241 -----+-----+-----+-----+-----+-----+-----+ 6300
      TTTTACTATTTTAGCCCTATGGTCCCATCACTCCTTTTCTACCAAGACAGTTTCATCTCA
e      L I I F I P I A P L S S F I T E T L Y L -
      GAGAAAAGGTACAGCCGATGCTGGTCGATAAATCCTTCAATCTTCTGCTCAGTCATCCAC
6301 -----+-----+-----+-----+-----+-----+-----+ 6360
      CTCTTTTCCATGTCCGCTACGACCTATTGAGGAAGTTAGAGACGAGTCAGTAGGTG
e      S F L Y L R H Q D I V G E I K Q E T M W -
```

266/487

Figure 1010

```
6361 TCTTGAACAATTGCTTTCGAAATATGATACAGTGGCTTGTGCTTTCAATCCCATATGT 6420
-----+-----+-----+-----+-----+-----+
AGAACTTGTTAACGAAAGCTTTATATACTATGTCACCGAACACGCGAAAGTTAGGGTATTACA
e E Q V I A K S I H Y L P K D S E I G Y H -

6421 TCGTAATAATTATAATAGGGAACCTAGATTTTGTAACCAACAAACGTTCTTGTTAAG 6480
-----+-----+-----+-----+-----+-----+
AGCATTATTAAATATCCCTTGATCTAAACATTTGGTTTGTGTTTTCGCAAGAACAAATTC
e E Y Y N Y Y P V L N Q L G F L F T R T L -

6481 AAAGTCAGTGTGTTAAAAAGAAAGAGAAATTCGAAATGTCATTTCCCTAAGATATCTTG 6540
-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAAATTTTCTCTTCTCTTAAGCTTTACAGTAAAGGATTTCTATAAGAAC
e F T L A T L F S L S N S I D N G L I N K -

6541 AACTGGATAGTAGTCTTCTCTTGTATGCTGAAGAAATCAGTTGAATAGTATGAGTC 6600
-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAGGAGAACATACGACTTCTTAGTCAACTTATCATACTCAG
e F K S L L H K G R T H Q L I L Q I T H T -

6601 TTTTCTTCTGATTCCTTCTGCTTGGAAACGAGAAATTAGCAGAACAAATAAACCAA 6660
-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAAACAGGAACCTTTTGTCTTCTTAATCGTCTGTTGTTTATTTGGTTT
e K K E Q N W K D K S F S S N A S C Y V L -

6661 AAGATATAATCCAGTTCTTCTGAGTAAAGTCATGTTGGCATGTGGCTCTAAGTAAGTT 6720
-----+-----+-----+-----+-----+-----+
TTCTATATTAGGTCAAGAAGGACTCATTTTCAGTACACCGTACACCGAGATTCATTCAA
e F I Y D L E E Q T F T M N A H P E L Y T -

6721 TGGCAATGTTCCATCAAAATCGGATACATAAAGAGGTTTTTTAATTTTCAAACTCTTTG 6780
-----+-----+-----+-----+-----+-----+
ACCGTTACAAGGTAGTTTTAGCCTATGTATTTCTCCAAAAAATTAAGTTTGAGAAAC
```

267/487

Figure 101P

e Q C H E M L I P Y M F L N K L K E F E K -
6781 GACTCAGGGAACCTCAAGTGGAAATCCCGACGTTTCCAGTGAGTGCCACTAGTATGCTA
CTGAGTCCCTTGAGTTCACCTTTAAGGCTGCAAGGTTCACTCAGGTGATCATACGAT 6840
e S E P F E L P F E R R K W T L A V L I S -
6841 AAATGAACATACTCGTCAGGTGTGATTTCTAACAGTTCATGACTGAGTTGAGAATTAGAC
TTTACTTGTATGAGCAGTCCACACTAAAGATTGTCAAGTACTGACTCAACTCTTAATCTG 6900
e F H V Y E D P T I E L L E H S L Q S N S -
6901 TGCACAATCATATGTGTGACCCCAATCCATACTTCCATCATTTCAAATCATAAATCTCAATA
ACGTGTTAGTATACACACTGGGTAGGTATGAAGGTAGTAAGTTAGTATTAGAGTTAT 6960
e Q V I M H T V W D M S G D N L D Y I E I -
6961 CCAAAATGAACCTGGAGAGTGCAATTAAAAACGAATGCGATATTCAGGACCAACTACT
GGTTTACTTTGACCTCCCTCAGGTTAATTTTGTGCTACGCTATAAGTCCCTGGTTGATGA
e G F H F Q L L A I L F R I R Y E P G V V -
7021 TGATTTTTCACAAGGTCCAAACCTACTGAACGTAGTAACAAGCCACACTTTTGTGCTGACG
ACTAAAAAGTGTCCAGGTTTGGATGACTTGCATCATTTGTTCCGTTGTGAAAAACAGCATGC 7080
e Q N K V L D L G V S R L L L G C K Q R V -
7081 CCGTAGCCCTGTTGCGATGGAAATATACTCTTTTGTGTAAATTCGTTAAAGCTTTGATTA
GCCATCGGACAAACGCTACCTTTATATGAGAAAAACACATTTAAGCAATTTGGAACATAAT 7140
e R Y G T A I S I Y E K Q T F E N F S Q N -
7141 CCTTGTAGTAAAGACGGAGTATTTTAAAAATAGTTGATTGTTTATAAAGCTGATGG
7200

268/487

Figure 101Q

GGAAACATCATCTTTCTTTGGCCCTCATTAAGAAATTTTATCAACTAACCAATATATTTCCGACTACC

ee G Q T F F R L I K L H T S Q N Y L Q H

7201 AAGTAAATAATTCGTTTGATGAGAAATGGTGTTCGATTAAATTGAACCTTGTTCGTATCTAAA 7260

TTTCATTATTAAAGCAAACTACTCTTACCCAGAAGCTAATTAACTTGAACAACGCATAGATTT

e
 F
 Y
 Y
 N
 T
 Q
 H
 S
 H
 H
 H
 I
 L
 Q
 V
 Q
 Q
 F
 D
 L

7261 TTAATGTCAACTCTTCCTCGAATGTTTCTTGTAAATTCCTGCATAAATGCTTAGGAGACTT 7320

AAATTTACAGTTGAGAGGAGCTTCAAAGAACATTAAAGGACGTTTACGAATCCTCTGAA

STILL QUESTIONS

7321
TTAGATTGTAATGAAGTTAAAGTAGACAGTTTCATCTAGTTCATTAGACCGAATATCCAAAT
7380

AAATCTAACATTACTTCAATTTCATCTGTCAGTAGATCAAGTTATCTGGCTTATAGTTA

EXQUISITELY DELICIOUS

7381 AATATATTAAAGGTAATTTTATCTCTGTAATCTTTTTCAATGTAATTGTTTAGCATA 7440

TTTATATAAATTTTACCATTAAATAGACATTAAAGAAAAGTTACATATAACAAATCGTAT

e
 l
 l
 n
 l
 l
 f
 f
 h
 h
 k
 k
 d
 t
 i
 r
 k
 e
 i
 y
 k
 n
 l
 m

7441 GTTACCGAATCTTAGTTGCATATAGATAAATTTTAAATTATTAATAACAAAGAACTAAT 7500

CAATGGCTTAGAATCAACGTATATCTATTAATAATAATAATATATGTTTCTTTTGATTAT

7501 TGTCTGTCAAAAAGGTTGTGGAAATTTCCGACTTTATTGATAAAACAGCATGTAATAAAA 7560

ACAGAACAGTTTTTCCAACACCTTAAAGCGTGAATAACTATTTTGTCCGTACATTATT

7561 GGCATTTTAAAGATAGTAAATGAGTATTGGTGGAGTTTTATGGCTATTTTTATTATAGA 7620

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PCT/US05/27239 269/487

Figure 101R

```
CCGTAAATTTCTATCATTAACCAACCTCAAAATACCGAATAAAAAAATAATCT
7621 -----+-----+-----+-----+-----+ 7680
AAATATTTTTTTATCAAAATATTGTCGTTCTATAAAAAATATGTGATAAAATATCTATT
TTTATAAAAAAATAGTTTATAACAGCAAGATATTTTTTTATACACTATTTTATAGATAA
7681 -----+-----+-----+-----+-----+ 7740
GTGATGGAAGTTGTTTTTAAATTTATACTAGGATAGTTAATAGTAATACTATACATAT
CACTACCTTCAACAAATTAATAATGATCCCTATCAATTATCATATGATATGATATGATA
7741 -----+-----+-----+-----+-----+ 7800
ATTGTATPACAAGTGTGTCATTGCCAGGTTGAGAAGATAGCTATAACGCACTTTATACGC
TAACATATGTTTACACACAGTAACGGTCCAACCTCTCTATCGATATTGCCGTGAAATATGCG
7801 -----+-----+-----+-----+-----+ 7860
TTTTTGCTACGTTTGTAGTGAACGGATTAACTCAGTGAGATAAATTTTATCAGAACATAA
AAAACGATGCAAAACAATCAGTTGCCCTAATTGAGTCACCTCTATTTAAATAGCTTTGTATT
7861 -----+-----+-----+-----+-----+ 7920
GTAATCCGTTTCTTCGTGTATACAGATTGAAAGTACCTATGAATCATAGAAGGATTAACT
CATTAGGCAAGAAGACACATATGTCCTAACTTTTCATGGATACCTTAGTATCTTCCCTAATTGA
7921 -----+-----+-----+-----+-----+ 7980
TGTTCTATGAATAATGCTTAACAGGGAGACACACATGAAAAAGTAAGAAAGATATTCA
ACAAGATACCTTATTACGAATTGTCCCTCTGTGTACTTTTTTTCATTTCTTCTATAAAGT
7981 -----+-----+-----+-----+-----+ 8040
GAAGGCAGTTGCAGGACTGTGTATATCTCAGTTGACAGCTTTTCTTCGATAGTTGC
CTTCCGTCAACGTCCTGACACGACATATAGAGTCAACTGTGCGAAAAAGAGCTATCAACG
b      M L N R E T H M K K V R K I F Q -orf3_670 homologue of sp0462, LPXTG
      K A V A G L C C I S Q L T A F S S I V A -
```

PCT/US05/27239

270/487

Figure 101S

8041 TTTAGCAGAAACGCCCTGAAACAGTCCAGCGATAGGAAAGTAGTGATTAAAGGAGACAGG + 8100
-----+-----+-----+-----+-----+-----+-----+
AAATCGTCTTTGCGGACTTTGGTCAGGTCGCTATCCTTTTCATCACAATAATTCCTCTGTCC
b L A E T P E T S P A I G K V V I K E T G -
CGAAGGAGGAGCGCTTCTAGGAGATGCCGCTTTTGAGTTGAAAAACAATACGGATGGCAC + 8160
-----+-----+-----+-----+-----+-----+-----+
GCTTCCTCCTCGGGAAGATCCTCTACGGCAGAACTCAACTTTTGTATTATGCCCTACCGTG
b E G G A L L G D A V F E L K N N T D G T -
AACTGTTTCGCAAGGACAGAGCGCGCAACAGGAGAGCGGATATTTCAACATAAAACC + 8220
-----+-----+-----+-----+-----+-----+-----+
TTGACAAAGCGTTTCCCTGTCTCCGCGTTTGTCTCTCGCTATAAAAGTTTGTATTGTTGG
b T V S Q R T E A Q T G E A I F S N I K P -
TGGGACATACACCTTGACAGAGCCCAACCTCCAGTTGGTTATAAACCCCTCTACTAAACA + 8280
-----+-----+-----+-----+-----+-----+-----+
ACCTGTATGTGGAACTGTCTTCGGGTTGGAGGTCAACCAATATTTGGGAGATGATTTGT
b G T Y T L T E A Q P P V G Y K P S T K Q -
ATGGACTGTTGAAGTTGAGAAAGATGGTCGGACGACTGTCCAAGGTGAACAGGTAGAAAA + 8340
-----+-----+-----+-----+-----+-----+-----+
TACCTGACAACTTCAACTCTTCTTACCAGCCCTGCTGACAGGTTCCACTTGTCCATCTTT
b W T V E V E K N G R T T V Q G E Q V E N -
TCGAGAAGAGCGCTCTATCTGACCAGTATCCACAACAGGGACTTATCCAGATGTTCAAAC + 8400
-----+-----+-----+-----+-----+-----+-----+
AGCTCTTCTCCGAGATAGACTGGTCAATAGGTGTTTGTCCCTGAATAGGTCTACAAAGTTTG
b R E E A L S D Q Y P Q T G T Y P D V Q T -
ACCTTATCAGATTATTAAAGGTAGATGGTTCGGAAAAACAGGACAGCACAGCGGTTGAA + 8460
-----+-----+-----+-----+-----+-----+-----+
TGGAAATAGTCTAATAATTCCATCTACCAAGCCCTTTTTCCTGCTGTTCCGCAACTT

PCT/US05/27239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -
8461 TCCGAATCCATATGAACGTGTGATTCAGAGGTACACTTTCAAAGAGAATTATCAAGT
-----+-----+-----+-----+-----+ 8520
AGGCTTAGGTACTTGCACACTAAGGCTTCCATGTGAAGTTTCTCTTAATAGTTCA
P N P Y E R V I P E G T L S K R I Y Q V -
8521 GAATAATTGGATGATAACCAATATGGAATCGAGTTGACGGTTAGTGTTAAACGACGGT
-----+-----+-----+-----+-----+ 8580
CTTATTAAACCTACTATTGGTTATACCTTAGCTCAACTGCCCAATCACCATTTTGTGCTGCCA
N N L D D N Q Y G I E L T V S G K T T V -
8581 TGAACGAAAGAGCCTCTACTCCGCTAGATGTTGTTATTCTATTAGATAACTCCAAATAG
-----+-----+-----+-----+-----+ 8640
ACTTTGCTTTCTTCGGAGATGAGGGGATCTACAACAATAAGATAATCTATTGAGGTTATC
E T K E A S T P L D V V I L L D N S N S -
8641 TATGAGTAATATTCGACATAATCATGCCATCGAGCGGAAAGCGGAGAGCGACACG
-----+-----+-----+-----+-----+ 8700
ATACTCATTATAAGCTGTATTAGTAGCGGTAGCTCGCCTTTTTCGCCCTCTTCGCTGTGC
M S N I R H N H A H R A E K A G E A T R -
8701 AGCCCTTGTAGATAAGATTACCTCCAAATCCAGATAATCGAGTAGCCTTGTGACTTATGG
-----+-----+-----+-----+-----+ 8760
TCGGGAACATCTATTCTAATGGAGGTTAGGTCTATTAGCTCATCGTGAACACACGTAATACC
A L V D K I T S N P D N R V A L V T Y G -
8761 CTCAACTATCTTTGACGGTTTCAGAAAGCTACTGTGGAAAAAGGGGTAGCAGATGCGAACGG
-----+-----+-----+-----+-----+ 8820
GAGTTGATAGAAACTGCCAAGTCTTCGATGACACCTTTTTCCTCCCATCGTCTACGCTTGCC
S T I F D G S E A T V E K G V A D A N G -
8821 AAAAATATTGAATGACTCAGCTTTATGGACGTTTCGATCGTACGACGTTTACAGCTAAAC
-----+-----+-----+-----+-----+ 8880

Figure 101U

TTTTTATACTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAAATGTCGATTTTG
K I L N D S A L W T F D R T T F T A K T -
TTATAATTATAGCTTTTAAATCTCACATCAGATCCTACTGATATTCAAACTATTAAAGGA 8940
-----+-----+-----+-----+-----+-----+
AATATTAAATATCGAAAAATTTAGAGTGTAGTCTAGGATGACTATAGTTTGATAATTCCT
Y N Y S F L N L T S D P T D I Q T I K D -
TAGGATTCATCAGATGCAGAGGAATGAACAAAGACAAATGATGATCAATTCGGCGC 9000
-----+-----+-----+-----+-----+-----+
ATCCTAAGGTAGTCTACGTCCTCCTTAACTTGTCTGTTTAACTACATAGTTAAGCCGCG
R I P S D A E E L N K D K L M Y Q F G A -
GACTTTTACCCAGAGGCTTTGATGACCGCTGATGATATCTTGACAAAGCAGGCAAGACC 9060
-----+-----+-----+-----+-----+-----+
CTGAAAAATGGGTCTTCCGAAACTACTGGCGACTACTATAGAACTGTTTCGTCCTCTGCG
T F T Q K A L M T A D D I L T K Q A R P -
AAACAGTAAAAAGGTTATTTTCCACATTACAGATGGTGTTCGACTATGTATATCCAAT 9120
-----+-----+-----+-----+-----+-----+
TTTGTCATTTTCCAAATAAAAGGTGTAATGCTTACCACAAAGGCTGATACAGTATAGGTTA
N S K K V I F H I T D G V P T M S Y P I -
TAATTTTAAATATACAGGAACGACGCAATCGTACAGAACTCAGCTGAATAATTTTAAAGC 9180
-----+-----+-----+-----+-----+-----+
ATTAAAAATTTATATGTCCTTGGCTGCTTAGCATGTCTTGAGTCGACTTATTAATAATTTTCG
N F K Y T G T T Q S Y R T Q L N N F K A -
AAAAACTCCAAATAGTAGCGGGATATTACTGGAGGACTTTGTTACATGGTCAGCAGATGG 9240
-----+-----+-----+-----+-----+-----+
TTTTTGAGGTTTATCATCGCCCTATATAATGACCTCCCTGAAACAATGTACCAGTCGTCCTACC
K T P N S S G I L L E D F V T W S A D G -

PCT/US05/27239

273/487

Figure 101V

9241 TGAACATAAGATTGTTCTGTTGGAGATGGTGAAGTTATCAGATGTTTACGAAGAAACCTGT 9300
-----+-----+-----+-----+-----+-----+-----+
ACTTGTATTCTAACAAGCACCTCTACCACTTTCAATAGTCTACAAATGCTTCTTTGGACA
b E H K I V R G D G E S Y Q M F T K K P V -
9301 AACAGACCAATACGGAGTTTCATCAAAATACTTTCAATCACCTCCATGGAGCAGAGAGCTAA 9360
-----+-----+-----+-----+-----+-----+-----+
TTGTCTGGTTATGCTTCAAGTAGTTTATGAAAAGTTAGTGGAGGTACCTCGTCTCTCGATT
b T D Q Y G V H Q I L S I T S M E Q R A K -
9361 ATTAGTTTTCAGCGGATATAGGTTCTATGGAAGTCTGACTTGTATTTATATTTGGCGTGATAG 9420
-----+-----+-----+-----+-----+-----+-----+
TAATCAAAGTGGCCCTATATCCAAAGATACCTTGACTGAACATAAATATAACCGCACATATC
b L V S A G Y R F Y G T D L Y L Y W R D S -
9421 TATTCTAGCCCTATCCATTTAACCTCTAGTACCGATTGGATTACCAACCATGGTGACCCCTAC 9480
-----+-----+-----+-----+-----+-----+-----+
ATAAGATCGGATAGGTAATGAGATCATGGCTAACCTAATGTTGGTACCACCTGGGATG
b I L A Y P F N S S T D W I T N H G D P T -
9481 GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGCTCTTCACTGTGGGGT 9540
-----+-----+-----+-----+-----+-----+-----+
CTGAACCATATAATTTGCCCTTTATACCGAGTCCCTACCGATACCTACAGAAAGTGACAACCCCA
b T W Y Y N G N M A Q D G Y D V F T V G V -
9541 TGGTGTAACGGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTATTCAGAGCAT 9600
-----+-----+-----+-----+-----+-----+-----+
ACCACATTTGCCCTTAGGACCATGCCCTACTTCTGTTGCCGATGATCTAAATACGTCCTCGTA
b G V N G D P G T D E A T A T R F M Q S I -
9601 CTCAGTTCTCTGACAACTACACTAAGTAGCAGATCCATCTCAGATTTTACAAGAATT 9660
-----+-----+-----+-----+-----+-----+-----+
GAGATCAAGAGGACTGTTGATGTGATTGCACTGCTAGGTAGAGTCTAAAATGTTCTTAA

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -
GAATCGCTACTTCTATATACTATCGTCAATGAGAGAGAAATCTATCGAAATGGTACGATTAC 9720
-----+-----+-----+-----+-----+
CTTAGCGATGAAGATATGATAGCAGTTACTCTTCTTTAGATAGCTTTTACCATGCTAATG
N R Y F Y T I V N E K K S I E N G T I T -
AGACCCGATGGGTGAACATAATTGATTTCCTCAATTGGGAGCAGATGGAAGGTTTGATCCAGC 9780
-----+-----+-----+-----+-----+
TCTGGGCTACCCACTTGATTAACTAAAGTTAACCCCTCGTCTACCTTCCAAACTAGGTCG
D P M G E L I D F Q L G A D G R F D P A -
GGATTACACTTTAACTGCAACGATGGTAGTTCTGTTGGTGAATAATGTCCCTACTGGGGG 9840
-----+-----+-----+-----+-----+
CCTAATGTGAAATTGACGTTTGCTACCATCAAGCAACCACCTATTACAGGGATGACCCCC
D Y T L T A N D G S S L V N N V P T G G -
ACCACAAAATGATGGTGGCTTGCTAAATAAAGTCAAAAGTGTCTATGATACGACTGAGAA 9900
-----+-----+-----+-----+-----+
TGGTGTTTTACTACCCGAAACGATTTTACGTTTTCACAAGATACTATGCTGACTCTT
P Q N D G G L L K N A K V F Y D T T E K -
AAGGATTGCTGTAACAGGTTTGTACCTTGGAAACGGGTGAAAAAGTTACATTGACTTATAA 9960
-----+-----+-----+-----+-----+
TTCCTAAGCACATTTGTCACCAACATGGAACCTTGCCCACTTTTTCATGTAAGTAATATT
R I R V T G L Y L G T G E K V T L T Y N -
TGTTCCGCTTGAATGACCAATTGTGAAGCAATAAATCTATGACACGAAATGGTCCGAACAAC 10020
-----+-----+-----+-----+-----+
ACAAGCGAACTTACTGGTTAAACATTCGTTATTAAAGATACTGTGCTTACCAGCTTGTG
V R L N D Q F V S N K F Y D T N G R T T -
CCTACACCCCTAAGGAAGTAGAAAGAACACAGTGGCGGACTTCCCGATTCCCTAAGATTCCG 10080
-----+-----+-----+-----+-----+
10021

Figure 101X

GGATGTGGGATTCCTTCATCTTTCTGTCACGGCTGAAGGCTAAGGATTCCTAAGC
b L H P K E V E K N T V R D F P I P K I R -
10081 TGATGTACGAAAGTATCCAGAAATCACAATTCCTCAAAAGAGAAAACTTGGTGAAATTGA
-----+-----+-----+-----+-----+ 10140
ACTACATGCTTTCATAGGTCCTTTAGTGTAAAGTTTCTCTTTTGAACCACTTTAACT
b D V R K Y P E I T I P K E K L G E I E -
10141 GTTTATTAAAGATCAATAAGAATGATAAAAAACCACCTGAGAGATGCGGCTTTAGTCTTCA
-----+-----+-----+-----+-----+ 10200
CAAAATAATTCTAGTTTCTTACTATTTTGGTGACTCTCTACGCCAGAAATCAGAAAGT
b F I K I N K N D K K P L R D A V F S L Q -
10201 AAAACAACATCCGGATTATCCAGATATTATGAGCTATTGATCAAAATGGCATTATCA
-----+-----+-----+-----+-----+ 10260
TTTTGTGTAGGCCCTAATAGGCTATAAATACCTCGATAACTAGTTTACCGTGAATAGT
b K Q H P D Y P D I Y G A I D Q N G T Y Q -
10261 AAAATGTGAGAACAGGTGAAGATGGTAAGTTGACCTTTAAAAATCTGTCAGATGGGAAATA
-----+-----+-----+-----+-----+ 10320
TTTACACTCTTGTCCACTTCTACCATTCACCTGGAATTTTGTAGACAGTCTACCCCTTTAT
b N V R T G E D G K L T F K N L S D G K Y -
10321 TCGATTATTGAAAAATCTGAACCAGCTGGTTATAAACCCGTTCAAAATAAGCCCTATCGT
-----+-----+-----+-----+-----+ 10380
AGCTAATAAACTTTAAGACTTGGTCGACCAATAATTTGGGCAAGTTTATTTCGGATAGCA
b R L F E N S E P A G Y K P V Q N K P I V -
10381 TGCCTTCCAAATAGTAAATGGAGAAAGTCAGAGATGTGACTTCAATCGTTCCACAAGATAT
-----+-----+-----+-----+-----+ 10440
ACGGAAGTTTATCATTTACCTCTTTCAGTCTCTACACTGAAGTTAGCAAGGTGTTCTATA
b A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

10441 ACCAGCGGGTTACGAGTTTACGAATGATAAGCACTATATACAAATGAGCCAATTCCTCC + 10500
-----+-----+-----+-----+-----+-----+-----+
TGGTCGCCCAATGCTCAATGCTTACTATTCGTGATATAGTGTCTTACTCGGTTAAGGAGG
b P A G Y E F T N D K H Y I T N E P I P P -
10501 AAAAAAGAGAAATATCCTCGAACTGGTGGTATCGGAATGTTCCTATCTGATAGGTTG + 10560
-----+-----+-----+-----+-----+-----+-----+
TTTTTCTCTTATAGGAGCTTGACCACCATAGCCCTTACAACGGTAAGATAGACTATCCAAC
K R E Y P R T G G I G M L P F Y L I G C -
b CATGATGATGGGAGGAGTTCTATTATATACACACGGAACATCCGTAAGTGTAGCAATGAG + 10620
-----+-----+-----+-----+-----+-----+-----+
GTACTACTACCCCTCAAGATAATATGTGTGCCCTTGTAGGCAATTCACATCGTTACTC
M M M G G V L L Y T R K H P *
10621 AAATGATAATATCGATPACTCTGAGCGATCTTTTAAAGTAGCACTCAAGAAGAGATTT + 10680
-----+-----+-----+-----+-----+-----+-----+
TTTACTATTATAGCTATGAGACTCGCTATGAAATCTTTCATCGTGAGTTCTCTCTAAA
AAGTTTACTTGGTGAAACAGTTTCTTCGCCAAGTAACACCATTGAAAGGGGAGATG + 10740
-----+-----+-----+-----+-----+-----+-----+
TTCAAATGAACCACCTTTGTCAAAAGAGCGGTTTCATTTGGTGGTAACTTTCCTCCTAC
TTTTTCGAAAACCTTGCACAGAAAAAGGATTATTATTGTCATGTGTAAATTCATTACATTGC + 10800
-----+-----+-----+-----+-----+-----+-----+
AAAAGCTTTTGAACGTGCTCTTTTTCCTAATAATAACAGTACACATTAAAGTAATGTAACG
TCACAGTTGATTTTAAGAGATATGAATAAGGAGAAATCATGAAATCAATCAACAAATTTT + 10860
-----+-----+-----+-----+-----+-----+-----+
AGTGTCAACTAAAAATCTCTATACTTATTCCTCTTTAGTACTTTAGTTAGTTGTTTAAA
M K S I N K F L - orf4_670, homologue of sp0463, LPXTG
c TAACAAATGCTTGCTGCCCTTATTACTGACAGCGAGTAGCCTGTTTTCAGCTGCAACAGTTT

277/487

Figure 101Z

```
10861 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 10920
      ATTGTTACGAACGACGGAATAATGACTGTGCTGCTCATCGGACAAAAGTCGACGTTGTCTAAA
c      T M L A A L L L T A S S L F S A A T V F -
      TTGCGGCGGACAAATGTTAGTACAGCACCAGATGCTGTACTAAAACCTTTAACAATCCATA 10980
10921 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      AACGCCGCCCTGTTACAATCATGTGCTGCTACGACAATGATTTTGAAATTTGTTAGGTAT
c      A A D N V S T A P D A V T K T L T I H K -
      AGTTACTGCTCTCAGAAAGATGATTTAAAGACTTGGGATACAAACGGTCCTAAAGGATATG 11040
10981 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TCAATGACGAGAGTCTTCTACTAAATTTCTGAACCTATGTTTGCCAGGATTTTCCCTATAC
c      L L L S E D D L L K T W D T N G P K G Y D -
      ATGGAACCTCAATCTAGTTTAAAGATTTAACTGGAGTTGTAGCTGAGGAAATTCCTCAAATG 11100
11041 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TACCTTGAGTTAGATCAAAATTTTCTAAATTGACCTCAACATCGACTCCTTTTAAGGTTTAC
c      G T Q S S L K D L T G V V A E E I P N V -
      TATACTTTGAATTACAAAAGTATAATTTGACTGATGGTAAGGAAAAGAAATCTTAAAG 11160
11101 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      ATATGAACCTTAATGTTTTCATATATTAACACTGACTACCATTCCTTTTCTTTTAGAATTC
c      Y F E L Q K Y N L T D G K E K E N L K D -
      ATGATAGTAAATGGACAAACAGTTTCATGGTGGTTTGACAACTAAAGATGGACTTAAATATG 11220
11161 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TACTATCATTTACCTGTTGTCAAGTACCACCAAACTGTGTGATTTCTACCTGAATTTTAAAC
c      D S K W T T V H G G L T T K D G L K I E -
      AAACCAGTACTCTTAAAGGTGTGTATCGTATTCGTGAGGATAGAACAAAGACTACCTATG 11280
11221 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TTTGGTCAATGAGAATTTCCACACATAGCATAAAGCACTCCCTATCTTGTGTTCTGATGGATAC
c      T S T L K G V Y R I R E D R T K T T Y V -
```

PCT/US05/27239

278/487

Figure 101AA

```
11281 TTGGTCCTAATGGGCAAGTATTAAACAGGTTCAAAGCCGTACCTGCTCTTGTAACTCTTC 11340
-----+-----+-----+-----+-----+-----+-----+
AACCAGGATTACCCGTTCAATAATTGTCCCAAGTTTTCGGCATGGACGAGAACATGAGAAG
C G P N G Q V L T G S K A V P A L V T L P -
11341 CACTTGTTAAACAATAATGGTACAGTAATTGATGCACATGTTTCCCTAAATAATCATATA 11400
-----+-----+-----+-----+-----+-----+-----+
GTGAACAATTGTTATTACCATGTTCATTAACACTACGTGTACAAAAGGGATTTTTAAAGTATAT
C L V N N N G T V I D A H V F P K N S Y N -
11401 ATAAACCAGTTGTAGATAAAAGAATTGCTGATACCTTTGAATTATAACGATCAAAATGGTC 11460
-----+-----+-----+-----+-----+-----+-----+
TATTGGTCAACATCTATTTTCTTAACGACTATGAAACTTAATAATTGCTAGTTTACCAG
C K P V V D K R I A D T L N Y N D Q N G L -
11461 TGTCTATCGGTACTAAAAATCCCATATGTTGTTAATAACAACAATTCCAAGTAATGCAACAT 11520
-----+-----+-----+-----+-----+-----+-----+
ACAGATAGCCATGATTTTAGGGTATACAACAATTATGTTGTTAAGGTTTCATTACGTTGTA
C S I G T K I P Y V V N T T I P S N A T F -
11521 TTGCAACTTCATTTTGGTCAGATGAATGACAGAGGCTCTAACTTATAATGAAGATGTAA 11580
-----+-----+-----+-----+-----+-----+-----+
AACGTTGAAGTAAACCCAGTCTACTTTACTGTCTTCCAGATTGAATATTACTTCTACATT
C A T S F W S D E M T E G L T Y N E D V T -
11581 CAATTACTTTTGAATAATGTAGCTATGGATCAAGCTGATTATGAAGTCACCTAAAGGAAATA 11640
-----+-----+-----+-----+-----+-----+-----+
GTTAATGAAACTTATTACATCGATACCTAGTTGGACTAATACTTCAGTGATTTCCTTTAT
C I T L N N V A M D Q A D Y E V T K G N N -
11641 ATGGCTTTAACTTAAAAATTAAACAGAAGCAGGTTTAGCTAAATAATGTTAAGGATGCAG 11700
-----+-----+-----+-----+-----+-----+-----+
TACCGAAATTGAATTTTAAATTGTCTTCGTCCTCCAAATCGATTTTAAATFACCATTCCTACGTC
```

C	G	F	N	L	K	L	T	E	A	G	L	A	K	I	N	G	K	D	A	D	-
11701	ACCAAAAAATCCAAATTA																				
	CTTACTCAGCTACTTTGAAC																				
	TCTGCTGTCAGACATTC																				
	TGGTTT																				
	TAAATGAATGAGTCGATGAAACTTGAGTGAACGACAACGCTCTGTAAG																				
C	Q	K	I	Q	I	T	Y	S	A	T	L	N	S	L	A	V	A	D	I	P	-
11761	CTGAAAGTAACGATATTACATATCATTAACGGAAATCATCAAGATCATGGGAATACTCCAA																				
	GACTTTCATTGCTATAATGTATAGTAATGCCCTTTAGTAGTTCTAGTACCCCTTATGAGGTT																				
C	E	S	N	D	I	T	Y	H	Y	G	N	H	Q	D	H	G	N	T	P	K	-
11821	AACCAACTAAACCTAAATAATGGTCAAAATTACAGTAAC																				
	TAAAGACATGGGACAGTCAACCTG																				
	TTGGTTGATTGGATTATTACCAGTTTAATGTCATTGATTCTGTTACCCCTGTCAGTTGGAC																				
C	P	T	K	P	N	N	G	Q	I	T	V	T	K	T	W	D	S	Q	P	A	-
11881	CTCCTGAGGGGTGAAGCGACTGTTCAACTTGTAAATGCCAAGACTGGTGAGAAAGTCG																				
	GAGGACTCCCCACTTTCGCTGACAAGTTGAACATTTACGGTCTGACCAC																				
	TCTTTCAGC																				
C	P	E	G	V	K	A	T	V	Q	L	V	N	A	K	T	G	E	K	V	G	-
11941	GTGCTCCTGTAGAACTTTCAGAAAAATAATGGACATATACTTGGAGTGGTCTAGATAAAT																				
	CACGAGGACATCTTGAAAGTCTTTTATTAACTGTATATGAACCTCACCAGATCTATTAA																				
C	A	P	V	E	L	S	E	N	N	W	T	Y	T	W	S	G	L	D	N	S	-
12001	CTATTGAATACAAAAGTTGAAGAAGAAATATAATGGATACTCAGCTGAATACACAGTAGAGA																				
	GATAACTTATGTTTCAACTTCTTCTTATATATACCTATGAGTCGACTTATGTGTCATCTCT																				
C	I	E	Y	K	V	E	E	E	Y	N	G	Y	S	A	E	Y	T	V	E	S	-
12061	GCAAAGGGAAGTTGGGGTAAAAAACTGGAAAGATAATAACCCAGCTCCAATCAATCCTG																				
	12120																				

PCT/US05/27239

280/487

Figure 101AC

```
CGTTTCCCTCAACCCCATTTTGTGACCTTCTATTTATTTGGTCGAGGTTAGTTAGGAC
      K G K L G V K N W K D N N P A P I N P E -
12121 AAGAACCCACGTTGTAATAACATACGGTAAAGTTTGTCAAGTAGACCACAAAAGATACTC
      +-----+-----+-----+-----+-----+
      TTCTTGGTGCACATTTTGTGTATGCCATTTTTCACACAGTTTCATCTGGTTTCTATGAG
      E P R V K T Y G K K F V K V D Q K D T R -
12181 GTCTAGAAAATGCCGAGTTCGTTGTTAAAAAGCAGATAGCAATAATATATGTCCTTTA
      +-----+-----+-----+-----+-----+
      CAGATCTTTTACGGCTCAAGCAACAATTTTTCGTCCTATCGTTATTTATATAACGGAAAT
      L E N A Q F V V K K A D S N K Y I A F K -
12241 AGTCAACTGCACAACAAGCTGCAGATGAAAAAGCAGCAGCAACTGCAAAAACAAAATTGG
      +-----+-----+-----+-----+-----+
      TCAGTTGACGTTGTTGTCGACGTCCTACTTTTTCGTCGTCGTTGACGTTTGTGTTTAAACC
      S T A Q Q A A D E K A A A T A K Q K L D -
12301 ATGCAGCGGTAGCAGCTTACACAATGCTGCAGATAAGCAAGCCGCTCAAGCTCTAGTAG
      +-----+-----+-----+-----+-----+
      TAGTGGCCCATCGTCGAATGTTTACGACGTCCTATTCGTTCCGGCAGTTCCGAGATCATC
      A A V A A Y T N A A D K Q A A Q A L V D -
12361 ATCAAGCACAGCAAGAATACAAATGTAGCTTACAAAGAACCAAAATTTGGTTATGTTGAAG
      +-----+-----+-----+-----+-----+
      TAGTTCCGTTGTCGTTCTTATGTTACATCGAATGTTTCTTCGGTTTAAACCAATACAACTTC
      Q A Q Q E Y N V A Y K E A K F G Y V E V -
12421 TAGCTGGAAGATGAAGCAATGGTTCTTACTTCTAATACGGATGGTCAATCCAAATTT
      +-----+-----+-----+-----+-----+
      ATCGACCTTTTCTACTTCGTTTACCAAGAATGAAGATTATGCCCTACCAGTTAAGGTTTAAA
      A G K D E A M V L T S N T D G Q F Q I S -
```

PCT/US05/27239 281/487

Figure 101AD

12481 CAGGTCCTGCTGGTACTTATAAATTAGAAGAAATTAAGCTCCAGAAGGTTTTCGGA + 12540
-----+-----+-----+-----+-----+-----+-----+
GTCCAGAACGACGACCATGAATATTAAATCTTCTTAATTTTCGAGGTCCTTCCAAAACGCT
C G L A A G T Y K L E E I K A P E G F A K -
12541 AAATTGATGATGATAATTGTTGTTGGAGCAGGTTCTTGGAAATCAAGGTGAGTTTAATT + 12600
-----+-----+-----+-----+-----+-----+-----+
TTTAACTACTACATCTTAAACAACAACCTCGTCCAGAACCCTTAGTTCCACTCAAATTA
C I D D V E F V V G A G S W N Q G E F N Y -
12601 ACTTAAAAGATGTTCAAAAAGATGACGCTACAAAAGTAGTCAACAAAAATCACTATCC + 12660
-----+-----+-----+-----+-----+-----+-----+
TGAATTTTCTACAAGTTTCTTACTGCGATGTTTTCATCAGTTGTTTTTTAGTGATAGG
L K D V Q K N D A T K V V N K K I T I P -
12661 CACAAACGGGTGATGTTGTAATAATCTTTGCTGAGCGGGGCTGCGATTATGGGTA + 12720
-----+-----+-----+-----+-----+-----+-----+
GTGTTTGCCACCATAAACCATGTTAATAGAAACGACATCGCCCCCGACGCTAATACCCAT
Q T G G I G T I I F A V A G A A I M G I -
12721 TTGCAGTGTAAGCATATGTTAAAAACAACAAGATGAGGATCAACTTGCTTAAGTAAGAG + 12780
-----+-----+-----+-----+-----+-----+-----+
AACGTCACATGCGGTATACAAATTTTGTGTTTCTACTCCTAGTTGAACGAATTCATTCTC
A V Y A Y V K N N K D E D Q L A *
12781 AGAAAGGAGCCATTGATGACAATGCAGAAAATGCAGAAAATGATTAGTCGTATCTCTTT + 12840
-----+-----+-----+-----+-----+-----+-----+
TCCTTCCCTCGGTAACACTACTGTTACGCTTTTACGCTCTTTTACTAATCAGCATAGAAGAA
a M T M Q K M Q K M I S R I F F -orf5_670, homologue of sp0464, LPXTG
12841 GTTATGGCTCTGTGTTTCTCTGTATGGGGTGCACATGCAGTCCAAAGCGCAAGAAGAT + 12900
-----+-----+-----+-----+-----+-----+-----+
CAATACCGAGACACAAAAGAGAACATACCCACCGTGTACGTACGTTTCGGTTCTCTCTA

PCT/US05/27239 282/487

Figure 101AE

```
a      V M A L C F S L V W G A H A V Q A Q E D -
      CACACGTTGGTCTTGCAATTGGAGAACTATCAGGAGGTGGTTAGTCAATTGCCATCTCGT
12901 -----+-----+-----+-----+-----+-----+-----+
      GTGTGCAACCAGAACGTTAACCTCTTGATAGTCCTCCACCACCAATCAGTTAACGGTAGAGCA
      H T L V L Q L E N Y Q E V V S Q L P S R -
      GATGGTCATCGGTTGCAAGTATGGAAGTTGGATGATTCGTATTCCTATGATGATCGGGTG
12961 -----+-----+-----+-----+-----+-----+-----+
      CTACCAGTAGCCCAACGTTTCATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCAC
      D G H R L Q V W K L D D S Y S Y D D R V -
      CAAATTGTAAGAGACTTGCAATTCGGTGGGATGAGAAATAAATTTCTTCTTCAAAAAGACT
13021 -----+-----+-----+-----+-----+-----+-----+
      GTTTAAACATTCCTGAACGTAAGCACCCCTACTCTTATTTGAAAGAGAAAGTTTTTCTGA
      Q I V R D L H S W D E N K L S S F K K T -
      TCGTTTGAGATGACCTTCCTTGAGAAATCAGATGAAGTATCTCATATCCAAATGGTCTT
13081 -----+-----+-----+-----+-----+-----+-----+
      AGCAAACTCTACTGGAAGGAACCTCTTAGTCTAACTTCATAGAGTATAAGGTTTACCAGAA
      S F E M T F L E N Q I E V S H I P N G L -
      TACTATGTTTCGCTCTATTATCCAGACGGATGCGGTTTCTTATCCAGCTGAATTTCTTTT
13141 -----+-----+-----+-----+-----+-----+-----+
      ATGATACAAGCGAGATAATAGGTCTGCCCTACGCCAAAGAAATAGGTCGACTTAAAGAAAAA
      Y Y V R S I I Q T D A V S Y P A E F L F -
      GAAATGACAGATCAAAACGGTAGAGCCCTTGGTCAATTGTAGCGAAAAAACAGATACAATG
13201 -----+-----+-----+-----+-----+-----+-----+
      CTTTACTGTCTAGTTTGCCATCTCGGAAACCAGTAACATCGCTTTTGTCTATGTATAC
      E M T D Q T V E P L V I V A K K T D T M -
      ACAACAAAGGTGAAGCTGATAAAGGTGGATCAAGACCACAAATCGCTTGGAGGGTGTCCGC
13261 -----+-----+-----+-----+-----+-----+-----+
      13320
```


PCT/US05/27239 283/487

Figure 101AF

```

TGTGTTTCCACTTCGACTAATTCACCTAGTCTGGTGTAGCGAACCTCCACAGCCG
a   T T K V K L I K V D Q D H N R L E G V G -
13321
TTTAAATTGGTATCAGTAGCAAGAGATGGTCTCTGAAAAGAGGTTCCCTTGATGGAGAA
-----+-----+-----+-----+-----+ 13380
AAATTAAACCATAGTCATCGTCTCTCTACCAAGACTTTTCTCTCCAAGGGAACCTAACCTCTT
F K L V S V A R D G S E K E V P L I G E -
a   TACCGTTACAGTCTCTCTGGTCAAGTAGGGAAGTCTCTATATACTGATAAAATGGAGAG
13381
ATGGCAATGTCAGAAGACAGTTCATCCCTCTCTGAGAGATAGTATTTTACCTCTC
Y R Y S S S G Q V G R T L Y T D K N G E -
a   ATTTTGTGACAAATCTCTCTCTGGGAACATACTGTTTCAAGGAGTGGAGCCACTGGCA
13441
TAAAACACACTGTTTAGAAGGAGAACCCTTGATAGCAAGTTCCTCCACCTCGGTGACCGT
I F V T N L P L G N Y R F K E V E P L A -
a   GGCTATGCTGTACGACGCTGGATACGGATGTCCAGCTGGTAGATCATCAGCTGGTGACG
13501
CCGATACGACAAATGCTGCGACCTATGCTACAGGTGCGACCATCTAGTAGTCGACCACTGC
G Y A V T T L D T D V Q L V D H Q L V T -
a   ATTACGGTTGTCAATCAGAAATTACCACGTTGGCAATGTTGACTTTATGAAGGTGGATGGT
13561
TAATGCCAACACAGTTAGTCTTTAATGGTGCACCGTTACAACCTGAAATACTTCCACCTACCA
I T V V N Q K L P R G N V D F M K V D G -
a   CGGACCAATACCTCTCTCAAGGGCAATGTTCAAGTCATGAAGAAGAAAGCGGACAC
13621
GCCTGGTTATGGAGAGAAGTTCCTCCGTTACAAGTTTCAGTACTTCTTCTTTCGCCCTGTG
R T N T S L Q G A M F K V M K E E S G H -

```

13681	a	TATACTCCTGTTCTTCAAAATGGTAAGGAAGTAGTTGTTACATCAGGGAAGATGGTCGT -----+-----+-----+-----+-----+-----+-----+ ATATGAGGACAAGAAGTTTACCATTCCTTCATCAACATGTAGTCCCTTTCTACCAGCA	13740
13741	a	Y T P V L Q N G K E V V V T S G K D G R - -----+-----+-----+-----+-----+-----+-----+ TTCCGAGTGGAGGCTAGAGTATGGGACATACTATTATTATGGGAGCTCCAAGCTCCAAC -----+-----+-----+-----+-----+-----+-----+ AAGGCTCACCTTCCAGATCTCATACCCCTGTATGATAAATAACCTCGAGGTTTCGAGGTGA	13800
13801	a	F R V E G L E Y G T Y Y L W E L Q A P T - -----+-----+-----+-----+-----+-----+-----+ GGTTATGTTCAATTAACATCGGCTGTTTCCCTTTACAATCGGGAAAAGATACTCGTAAGGAA -----+-----+-----+-----+-----+-----+-----+ CCAATACAAGTTAATTGTAGCGGACAAAGGAAATGTTAGCCCTTTCTATGAGCATTCCTT	13860
13861	a	G Y V Q L T S P V S F T I G K D T R K E - -----+-----+-----+-----+-----+-----+-----+ CTGGTAACAGTGGTTAAAAATAACAAGCCACCGGATTGATGTGCCAGATACAGGGGAA -----+-----+-----+-----+-----+-----+-----+ GACCATGTGCACCAATTTTATTGTTGCTGGTGCCCTAACTACACGGTCTATGTCCCCCTT	13920
13921	a	L V T V V K N N K R P R I D V P D T G E - -----+-----+-----+-----+-----+-----+-----+ GAAACCTTGATATCTTGATGCTTGTGTCATTTTGTGTTGGTAGTGGTTATTATCTT -----+-----+-----+-----+-----+-----+-----+ CTTTGGAACATATAGAAGTACGAACAACGGTAAACAACAACCAATCACCATAATAGAA	13980
13981	a	E T L Y I L M L V A I L L F G S G Y Y L - -----+-----+-----+-----+-----+-----+-----+ ACGAAAAACCAATAACTGATATTCAATGTACATCATTAATGAAAAAGATAGCAGGCTGA -----+-----+-----+-----+-----+-----+-----+ TGCCTTTTGGTTTATTGACTATAAGTTACATGTAGTAATACTTTTCTATCGTCCGACT	14040
14041	a	T K K P N N * -----+-----+-----+-----+-----+-----+-----+ AGGGAAGACCAGAGTACTCTGAGGTGATGTTAATCAGGAATCATGGTGATGTGGCATGAA -----+-----+-----+-----+-----+-----+-----+ TCCCTTCTGGTCTCATGAGACTCCACTACAATTAGTCTTAGTACCCTACACCGTACTT	14100

Figure 101AH

```
14101 TCACAAATACGGATATGAGGCTGGGAGATTGTGCCAGCCCTCATTTGTGGGTTATTGTTTG 14160
-----+-----+-----+-----+-----+-----+-----+
14161 AGTGTTATTGCCCTATACTCCGACCCGCTCTAACACGGTCGGAGTAACACCCCAATAACAAC
-----+-----+-----+-----+-----+-----+-----+

14161 TAAACGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGGACTGGGATTCGTATT 14220
-----+-----+-----+-----+-----+-----+-----+
14161 ATTTTGCTATCCTGACGACCATTAAGTAAATCCTTACCTGTCTGACCCCTAAGACTAA
-----+-----+-----+-----+-----+-----+-----+

14221 TAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT 14280
-----+-----+-----+-----+-----+-----+-----+
14221 ATTTTACCTACCACCTTAGTCTTTCTTTTACTCTCTAAAGAGCAAGAGAAATCGTCTATCCCTA
-----+-----+-----+-----+-----+-----+-----+

14281 TGTCTGTAGGAAAGCGATAAAATGATGAGTTTGAAGATAAAGGATGCTGATAAAAT 14340
-----+-----+-----+-----+-----+-----+-----+
14281 ACAGACAATCCTTTTCCGCTATTTTACTACTCAAACTTCTATTTCCCTACGACTATTTTA
-----+-----+-----+-----+-----+-----+-----+

14341 GGTAACAAACAAAGCAAAACGAAATAATCTCCTATTAGGAGTGGTATTTTTCATTGG 14400
-----+-----+-----+-----+-----+-----+-----+
14341 CCATTTTGTGTTTTTTCGTTTTTTCGTTTATTAGAGGATAATCTCCTCACCATAAAAGTAACC
-----+-----+-----+-----+-----+-----+-----+

b V K T K K Q K R N N L L L G V V F F I G -
14401 AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGCTTGTTATTCGAGTGGAAATCAATCA 14460
-----+-----+-----+-----+-----+-----+-----+
14401 TTACCGCCCATACCGCATAGGCGACACAGAGCGAACAATAATAGCTCACCTTAGTTTAGT
-----+-----+-----+-----+-----+-----+-----+

b M A V M A Y P L V S R L Y Y R V E S N Q -
14461 ACAAAATTGCTGACTTTGATAAGGAAAGCAACGTTGGATGAGGCTGACATTTGATGAACG 14520
-----+-----+-----+-----+-----+-----+-----+
14461 TGTTTAACGACTGAAACTATTCTTTTTCGTTGCAACCTACTCCGACTGTAACTACTTGC
-----+-----+-----+-----+-----+-----+-----+

b Q I A D F D K E K A T L D E A D I D E R -
AATGAAATTGGCACAAAGCCTTCAATGACTCTTTTGAATAATGTAGTGAGTGGCGATCCTTG
```

M L I K M -orf6_670, homologue of sp0466, sortase

PCT/US05/27239

286/487

Figure 101A1

14521 -----+-----+-----+-----+-----+-----+-----+ 14580
TTACTTTAACCGTGTTCGGAAGTTACTAGAAACTTATACATCACTACCGCTAGGAAC
b M K L A Q A F N D S L N N V V S G D P W -
GTCCGAAGAAATGAAGAAAAAGGGCGAGAGATATGCACGTATGTAGAAATCCATGA 14640
14581 -----+-----+-----+-----+-----+-----+-----+
CAGCCTTCTTTTACTTCTTTTTCCTCCGCTCTCATACGTGCATACAACTTTAGGTACT
b S E E M K K K G R A E Y A R M L E I H E -
GGGATGGGGCATGTGGAAATCCCGTTATTGACGTGGATTGCCGGTTTATGCTGGTAC 14700
14641 -----+-----+-----+-----+-----+-----+-----+
CGCCTACCCCGTACACCTTTAGGGGCAATAAAGTGCACCTAAACGGCCAAATACGACCATG
b R M G H V E I P V I D V D L P V Y A G T -
TGCTGAAGAGGTATTGCAGCAAGGGGCTGGGCATCTAGAGGGAACCTCTCTGCCGATCGG 14760
14701 -----+-----+-----+-----+-----+-----+-----+
ACGACTTCTCCATAACGTCTGTTCCCGACCCCGTAGATCTCCCTTGAAGAGACGGCTAGCC
b A E E V L Q Q G A G H L E G T S L P I G -
AGGCAATTGACCCCATGGGTGATTACGGGCACATACAGGTTTGCCACAGCTAAGATGTT 14820
14761 -----+-----+-----+-----+-----+-----+-----+
TCCGTTAAGCTGGGTACGCCCACTAATGCCGTGTATGTCCAAACGGTTGTGATTTCTACAA
b G N S T H A V I T A H T G L P T A K M F -
TACGGATTGACCAAACTTAAAGTTGGGATAAGTTTATGTGCACAATATCAAGGAAGT 14880
14821 -----+-----+-----+-----+-----+-----+-----+
ATGCCATAAAGTGGTTGAATTTCAACCCCTATTCAAAATACACGTGTATAGTTCCCTTCA
b T D L T K L K V G D K F Y V H N I K E V -
GATGGCCTATCAAGTGGATCAAGTAAAGGTATTGAGCCGACGAACCTTGTATGATTTATT 14940
14881 -----+-----+-----+-----+-----+-----+-----+
CTACCGGATAGTTCACCTAGTTTCATTTCCACTAATCGGCTGCTTGAACACTACTAAATAA
b M A Y Q V D Q V K V I E P T N F D D L L -

$$287/487$$

14941	b	GATTGTACCAGGTCATGATTATGTGACCTTGCTGACTTGTTACGCCATACATGATCAATAC -----+-----+-----+-----+-----+-----+-----+ CTAACATGGTCCAGTACTAATACACTGGAACGACTGAACATGCGGTATGTACTAGTTATG -----+-----+-----+-----+-----+-----+-----+ 15000
15001	b	I V P G H D Y V T L L T C T P Y M I N T - -----+-----+-----+-----+-----+-----+-----+ CCATCGTCTATTGGTTCGGGGGCATCGGATACCGTACCGTAGCAGAGGTTGAGGAAGAATT -----+-----+-----+-----+-----+-----+-----+ 15060 GGTAGCAGATAACCAAGCCCCGGTAGCCTATGGCATCGTCTCCAACCTCCTTCTTAA
15061	b	H R L L V R G H R I P Y V A E V E E F - -----+-----+-----+-----+-----+-----+-----+ TATTGCAGCAAACTCAGTCATCTCTATCGCTACCTGTTTATGTGGCAGTTGGTTT -----+-----+-----+-----+-----+-----+-----+ 15120 ATAACGTCGTTGTTGAGTCAGTAGAGATAGCGATGGACAAAATACACCGTCAACCAAA
15121	b	I A A N K L S H L Y R Y L F Y V A V G L - -----+-----+-----+-----+-----+-----+-----+ GATTGTGATCTTTTATGGATTATTTCGACGCTTGCGCAAGAAAGAAAAACAACCGGAAAA -----+-----+-----+-----+-----+-----+-----+ 15180 CTAACACTAAGAAAAATACCTAATAAGCTGCGAACGCGTTCTTCTTTTGTGGCCTTTT
15181	b	I V I L L W I I R R L R K K K Q P E K - -----+-----+-----+-----+-----+-----+-----+ GGCTTTGAAGGCGCTGAAAAGCAGCAAGGAAGGAAGTGAAGGTGGAGGATGGACAACAGTA -----+-----+-----+-----+-----+-----+-----+ 15240 CCGAAACTTCCGGCACTTTCGTGCTTCCTTCCTTCACTTCCACCTCCTACCTGTGTGTCAT
15241	b	A L K A L K A A R K E V K V E D G Q Q * - -----+-----+-----+-----+-----+-----+-----+ GACGTTACGAAAAAAGGCACAAAAAAGAAACATCCGCTGATCCTTCTTCTGATTT -----+-----+-----+-----+-----+-----+-----+ 15300 CTGCAAGTGCCTTTTTCGGTGTCTTTTCTTCTTCTTTGTAGCGCACTAGGAAGAAGACTAAA
15301	b	TCTTAGTAGGATTCGCCGTTGGGATATATCCATTGGTGTCTCGTTATTATTATCGTATTG -----+-----+-----+-----+-----+-----+-----+ 15360 AGAATCATCTTAAGCGGCAACGCTATATAGGTAAACCACAGAGCAATAATAATAGCATTAAC

V S R Y Y R I E -orf7_670, homologue of sp0467, sortase

PCT/US05/27239

288/487

Figure 101AK

15361 AGTCAAACGAGGTTATTAAGAGTTTGATGAGACGGTTTCCAGATGGATAAGCAGAAC 15420
-----+-----+-----+-----+-----+-----+-----+
TCAGTTTGCTCCAAATAATTCTCAAACTACTCTGCCAAAGGTTCTACCTATTCCGTCCTTG
S N E V I K E F D E T V S Q M D K A E L -
15421 TTGAGGAGCGTTGGCGCTTGCGTCAAGCCCTTCAATGCGACCTTGAAACCATCTGAAATTC 15480
-----+-----+-----+-----+-----+-----+-----+
AACTCCTCGCAACCGCAACCGAGTTCGGAAGTTACGCTGGAACCTTTGGTAGACTTTAAG
E E R W R L A Q A F N A T L K P S E I L -
15481 TTGATCCCTTTTACAGAGCAAGAGAAAAAGAGCGTCTCAGAAATATGCCAATATGCTAA 15540
-----+-----+-----+-----+-----+-----+-----+
AACTAGGAAATGTCGTTCTCTTTTCTTTTCTTCCGACAGTCTTTATACGGTTATACGATT
D P F T E Q E K K K G V S E Y A N M L K -
15541 AGGTCCATGAGCGGATTGGCTATGTGGAATTCCTCGGATTGATCAGGAAATCCGATGT 15600
-----+-----+-----+-----+-----+-----+-----+
TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAAGTCTTAAAGGCTACA
V H E R I G Y V E I P A I D Q E I P M Y -
15601 ATGTCGGAACGAGTGAGGAAATCTTCTCAGAAAGGCGCAGGATTGCTAGAGGGAGCTTCGT 15660
-----+-----+-----+-----+-----+-----+-----+
TACAGCCTTGCTCACTCCTTTAAGAAAGTCTTCCCGGCTCCTAACGATCTCCCTCGAAGCA
V G T S E E I L Q K G A G L L E G A S L -
15661 TACCGGTTGGTGTAATAATACCCACACAGTTGTCTACGTCTCATAGAGGATTACCGACGG 15720
-----+-----+-----+-----+-----+-----+-----+
ATGGCCCAACCACTTTTATGGGTGTGTCAACAGTGACGAGTATCTCCTAATGGCTGCC
P V G G E N T H T V V T A H R G L P T A -
15721 CAGAACTGTTTAGTCAATTGGATAAGATGAAAAAGGGATGTCTTTTATCTTCACGTTT 15780
-----+-----+-----+-----+-----+-----+-----+
GTCTTGACAAATCAGTTAACCTATTCTACTTTTTTCCCTTACAGAAATAGAAAGTGCAAA

PCT/US05/27239

289/487

Figure 101A

c E L F S Q L D K M K K G D V F Y L H V L -
TAGACCAGGTGTTGGCCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG 15781
ATCTGGTCCACAACCGGATGGTTACACCTAGTCTAAAACTGCCAACTCGGTTTACTGAAAC 15840
D Q V L A Y Q V D Q I L T V E P N D F E -
AGCCTGCTTGATTCAACATGGGGAAGATTATGCGACCTTGTGACCTGTACACCGTATA 15841
TCGGACAGAACTAAGTTGTACCCCTTCTAATACGCTGGAACAACACTGGACATGTGGCATAT 15900
P V L I Q H G E D Y A T L L T C T P Y M -
TGATTAAACAGTCATCGTCTGTTGGTACGTGGGAAGCGGATTCCGTATACGGCACCAATTG 15901
ACTAATGTGTAGTAGCAGACAACCAATGCACCCCTTCGCCCTAAGGCATATGCCGTGGTTAAC 15960
I N S H R L L V R G K R I P Y T A P I A -
CAGAGCGAAATCGAGCGGTGAGAGAGCGTGGGCAATTCTGCTGTTGTTATTGCTAGCGG 15961
GTCTCGCTTTAGCTCGCCACTCTCTCGCACCCGTTAAGACCAACACCAATAACGATCGCC 16020
E R N R A V R E R G Q F W L L L A A -
CGTTGGTTATGATTCTGGTATTGAGTTACGGGGTGTATCGTCATCGTCGCATTGTCAAAG 16021
GCAACCAATACTAAGACCATAACTCAATGCCCCACATAGCAGTAGCAGCGTAACAGTTTC
L V M I L V L S Y G V Y R H R R I V K G -
GGCTAGAAAACAATTGGAGGAGCATCATGTCAAAGGCTAAGCTACAGAAATTACTAGGG 16081
CCGATCTTTTGTAACTCGCTCGTAGTACAGTTTCCGATTCCGATGCTCTTAATGATCCC 16140
a M S K A K L Q K L L G -orf8_670, homologue of sp0468, sortase
c L E K Q L E E H H V K G *

PCT/US05/27239

290/487

Figure 101AM

```
16141 TATTGCTGATGCTGTAGCATGGTGATTCCTGTTTATTGTTTGGGCGAGATGGTGTTA 16200
-----+-----+-----+-----+-----+-----+
16141 ATAAACGACTACGACCATCGTAACCACTAAGGACAAATAACAAACCCGCTTACCACAAT
-----+-----+-----+-----+-----+-----+
a Y L L M L V A L V I P V Y C F G Q M V L -
-----+-----+-----+-----+-----+-----+
16201 CAGTCTTTAGGACAAGTAAAGGTCATGAGATATTTTCAGAAATCTGTGACGGCCGACAGT 16260
-----+-----+-----+-----+-----+-----+
16201 GTCAGAAATCCTGTTTCATTTTCCAGTACTCTATAAAAGTCTTAGACACTGCCGGCTGTCA
-----+-----+-----+-----+-----+-----+
a Q S L G Q V K G H E I F S E S V T A D S -
-----+-----+-----+-----+-----+-----+
16261 TACCAAGAGCAATTGCAACGGTCGCTTGATTACAATCAACGCTTGATTGCAAAATCGT 16320
-----+-----+-----+-----+-----+-----+
16261 ATGGTTCTCGTTAACGTTGCCAGCGAACAATAATGTAGTTGCGAACCTAAGCGTTTATGCA
-----+-----+-----+-----+-----+-----+
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -
-----+-----+-----+-----+-----+-----+
16321 ATTGTAGATCCCTTTTGGCGGAAGGGTATGAGGTAAATTACCAAGTGTCTGACGATCCT 16380
-----+-----+-----+-----+-----+-----+
16321 TAACATCTAGGAAAAACCGCCTTCCCATCTCCATTAAATGGTTCACAGACTGCTAGGA
-----+-----+-----+-----+-----+-----+
a I V D P F L A E G Y E V N Y Q V S D D P -
-----+-----+-----+-----+-----+-----+
16381 GATGCAGTCTACGGCTATTGTGCGATTCCGAGTTTGGAATCATGGAGCCAGTTTATCTA 16440
-----+-----+-----+-----+-----+-----+
16381 CTACGTCAGATGCCGATAAACACAGCTAAGGCTCAACCTTTAGTACCTCGGTCAATAGAT
-----+-----+-----+-----+-----+-----+
a D A V Y G Y L S I P S L E I M E P V Y L -
-----+-----+-----+-----+-----+-----+
16441 GGAGCGGATTACCATCTATTAGCAATGGGGTTGGCCCATGTGGATGGGACGCCCTCTTCCT 16500
-----+-----+-----+-----+-----+-----+
16441 CCTCGCCTAATGGTAGTAAATCGTTACCCCAACCGGGTACACCTACCTGCGGAGAAGGA
-----+-----+-----+-----+-----+-----+
a G A D Y H H L A M G L A H V D G T P L P -
-----+-----+-----+-----+-----+-----+
16501 GTTGAGGGAAGGATTCGTTTCAGTGATTCCTGGCCACCGTGCAGAACCAAGCCATGTC 16560
-----+-----+-----+-----+-----+-----+
16501 CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCCGTGGCAGCTCTTGGTTCGGTACAG
```

Figure 101AM

a V E G K G I R S V I A G H R A E P S H V -
TTTTTCCGCCATTGGATCAGCTAAAGTTGGAGATGCTCTTTATTATGATAATGGCCAG 16561
-----+-----+-----+-----+-----+-----+-----+
AAAAAGCGGTAACCTAGTCGATTTTCAACCTCTACGAGAAATAATACTATTACCGGTC 16620
F F R H L D Q L K V G D A L Y Y D N G Q -
GAAATTGTAGAAATATCAGATGATGGACACAGAGATTATTTACCGTCGGAATGGGAAAAA 16621
-----+-----+-----+-----+-----+-----+-----+
CTTTAACATCTTATAGTCTACTACCTGTGTCTCTAATAAAATGGCAGCCTTACCCCTTTT 16680
E I V E Y Q M M D T E I I L P S E W E K -
TTAGAAATCGGTTAGCTCTAAAAATATCATGACCTTGATAACCTGGGATCCGATTCCTACC 16681
-----+-----+-----+-----+-----+-----+-----+
AATCTTAGCCAAATCGAGATTTTATAGTACTGGAACCTATTGGACGCTAGGCTAAGGATGG 16740
L E S V S S K N I M T L I T C D P I P T -
TTTAATAAACGCTTATTAGTGAATTTTGAACGAGTCGCTGTTATCAAAAATCAGATCCA 16741
-----+-----+-----+-----+-----+-----+-----+
AAATATTGCGAATAATCACTTAAACTTGTCTCAGCGACAAATAGTTTATTAGTCTAGGT 16800
F N K R L L V N F E R V A V Y Q K S D P -
CAAACAGCTGCAGTTGCGAGGTTGCTTTTACGAAAGAGACAATCTGTATCGCGTGT 16801
-----+-----+-----+-----+-----+-----+-----+
GTTTGTGACGTCACGCTCCCAACGAAATGCTTCTCTGTTAGACATAGCGCACAA 16860
Q T A A V A R V A F T K E G Q S V S R V -
GCAACCTCTCAATGGTTGTACCGTGGGCTAGTGGTACTGGCATTTCTGGGAATCCGTGTT 16861
-----+-----+-----+-----+-----+-----+-----+
CGTTGGAGAGTTACCAACATGGCACCCTGATCACCATGACCGTAAAGACCCCTTAGGACAAA 16920
A T S Q W L Y R G L V L A F L G I L F -
GTTTGTGGAAGCTAGCACGTTTACTACGAGGGAATAAAAGAAATGAAAGGAAAGCTA 16921
-----+-----+-----+-----+-----+-----+-----+ 16980

PCT/US05/27239 292/487

Figure 101AN

CAAAACACCTTCGATCGTGCAAAATGATGCTCCCTTTATTTTCTTTTACTTTCCTTTTCGAT

V L W K L A R L L R G K *

AGGCTGTTCCTTTTCCGGCTCTTTTGCAACTGTAGTGGGTGAAAAAAGCTAAGCTCG

16981 -----+-----+-----+-----+-----+-----+-----+ 17040

TCCGACAAAGGAAAAAGGCCGAGAAACAGTTGACATCACCCAACTTTTTCGATTTCGAGC

AGAAAGGACAAATTTTGTCCTTTCTTTTGTATATTCAGAGCGATAAAATCCGTTT

17041 -----+-----+-----+-----+-----+-----+-----+ 17100

TCTTTCCTGTTTAAACAGGAAAGAAAAAACTATAAGTCTCGCTATTTTATAGGCAAAAAA

GAAAGTTTTCAAA

17101 -----+--- 17112

CTTCAAAAAGTTT

PCT/US05/27239

293/487

M1, strain 2580

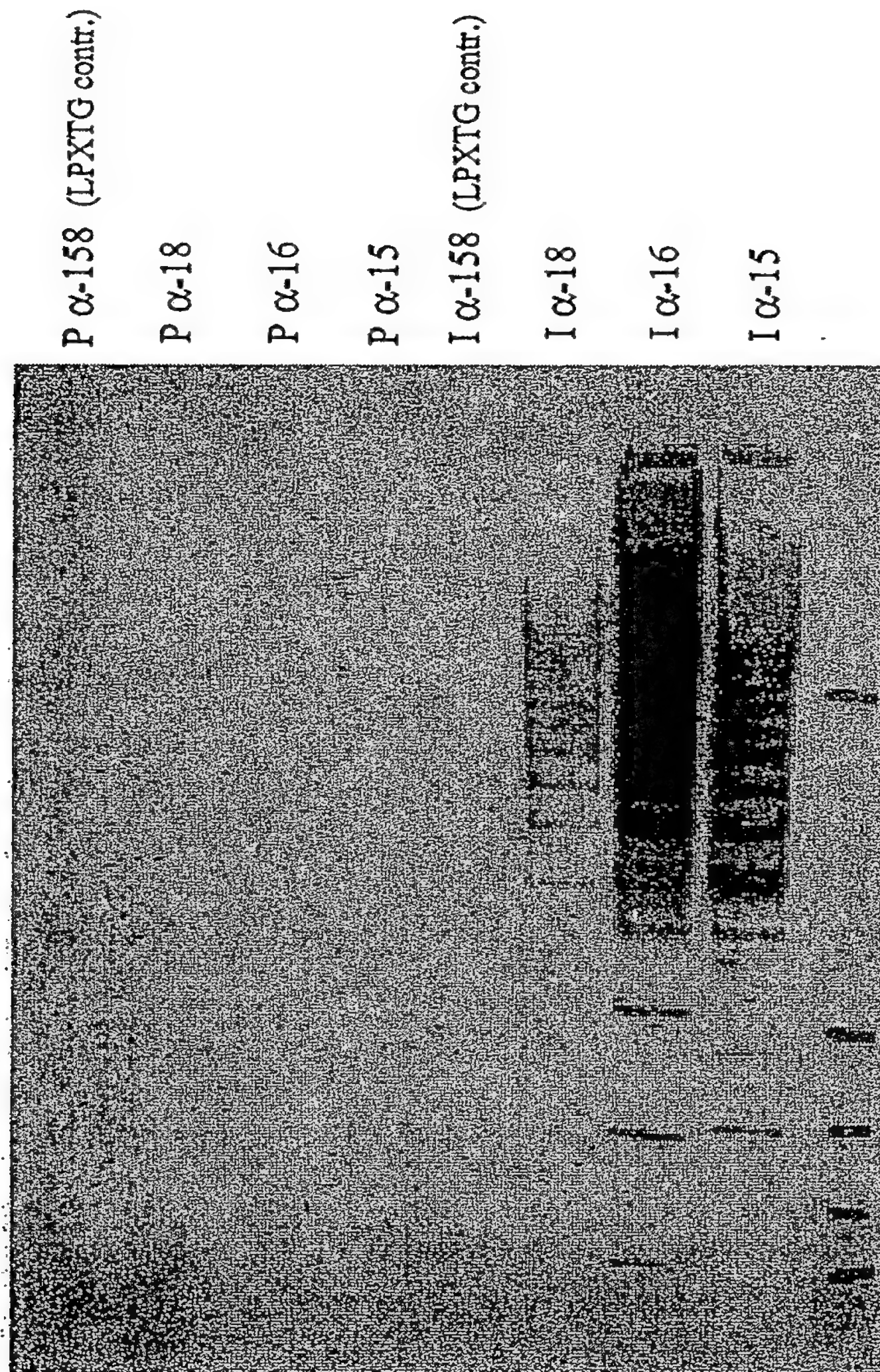


Figure 102

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

PCT/US2005/027239 294/487

M1, strain 2913

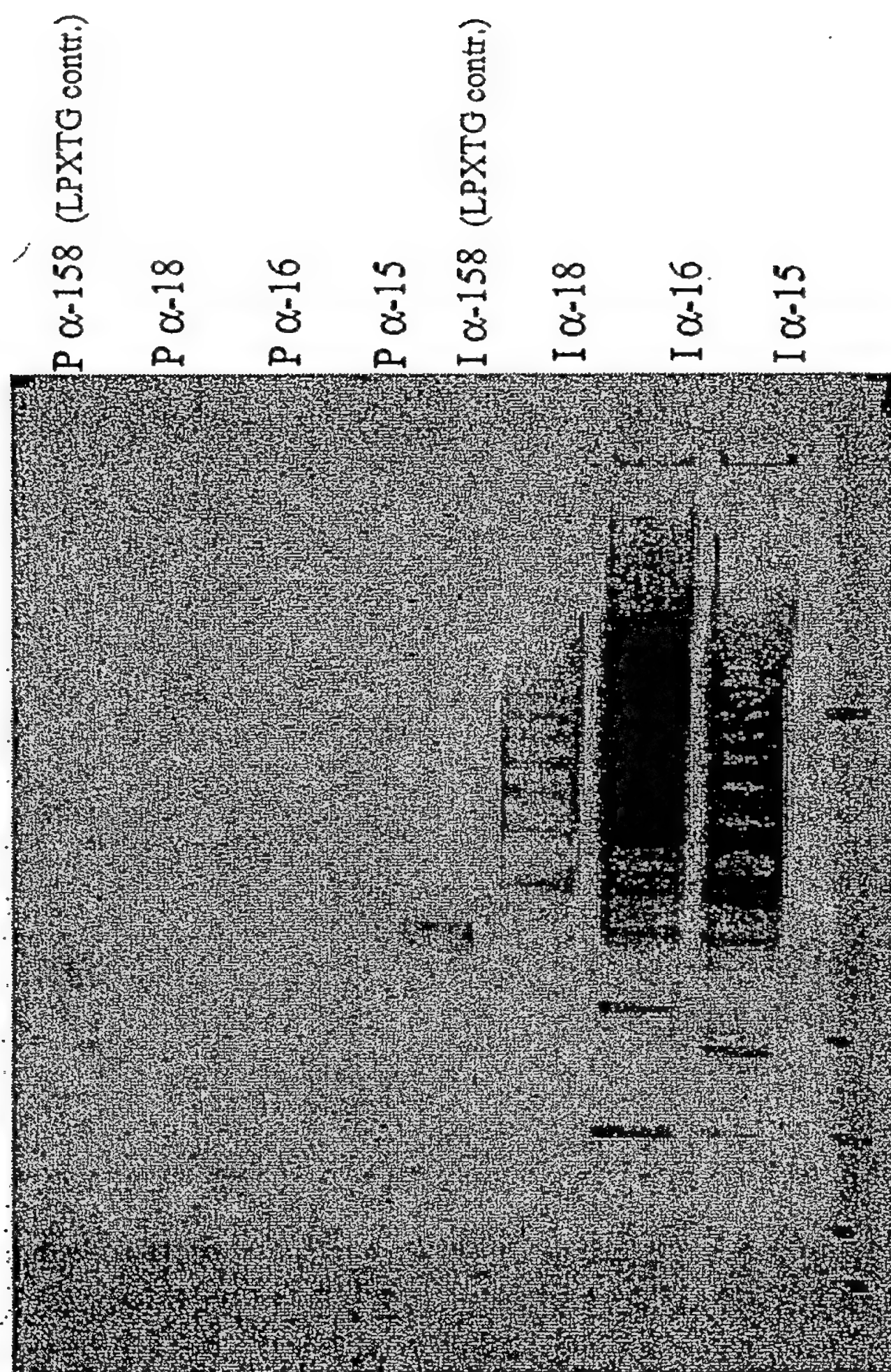


Figure 103

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

295/487

M1, strain 3280

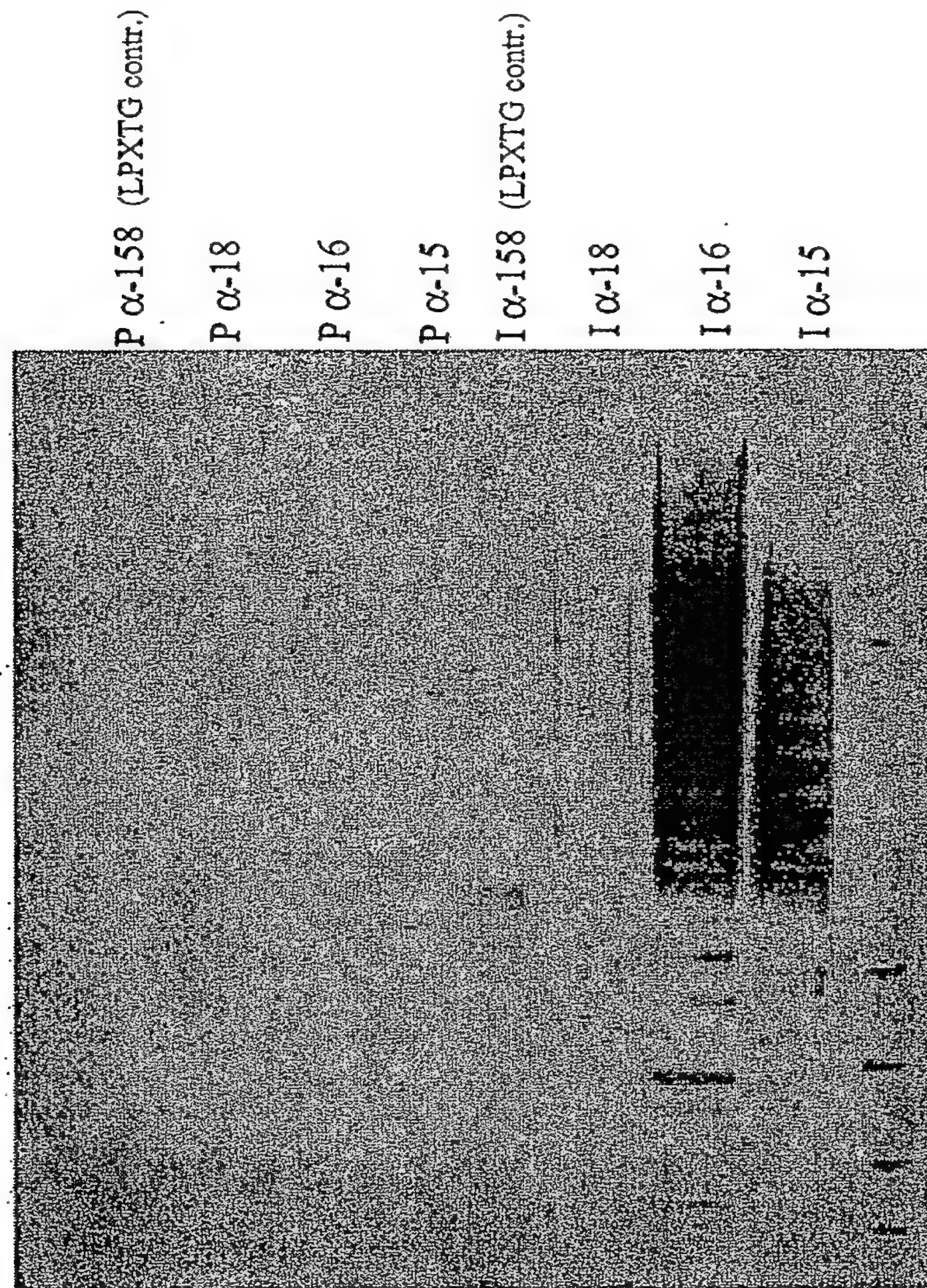
**LEGEND:**I α -#: immune serum anti-#P α -#: pre-immune serum anti-#

Figure 104

296/487

M1 strain 3348

P α -158 (LPXTG contr.)
P α -18
P α -16
P α -15
I α -158 (LPXTG contr.)
I α -18
I α -16
I α -15

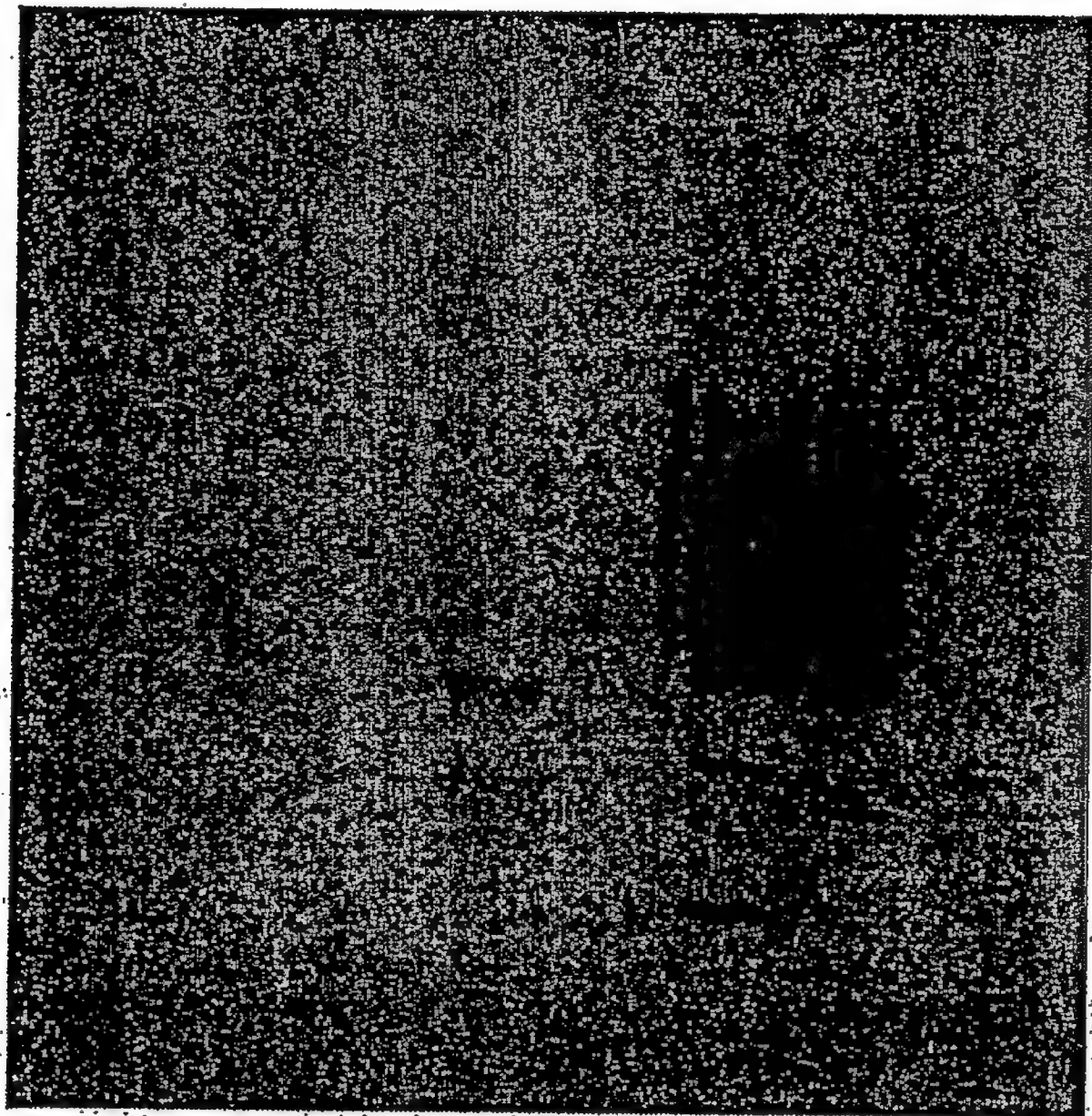


Figure 105

M1 strain 2719

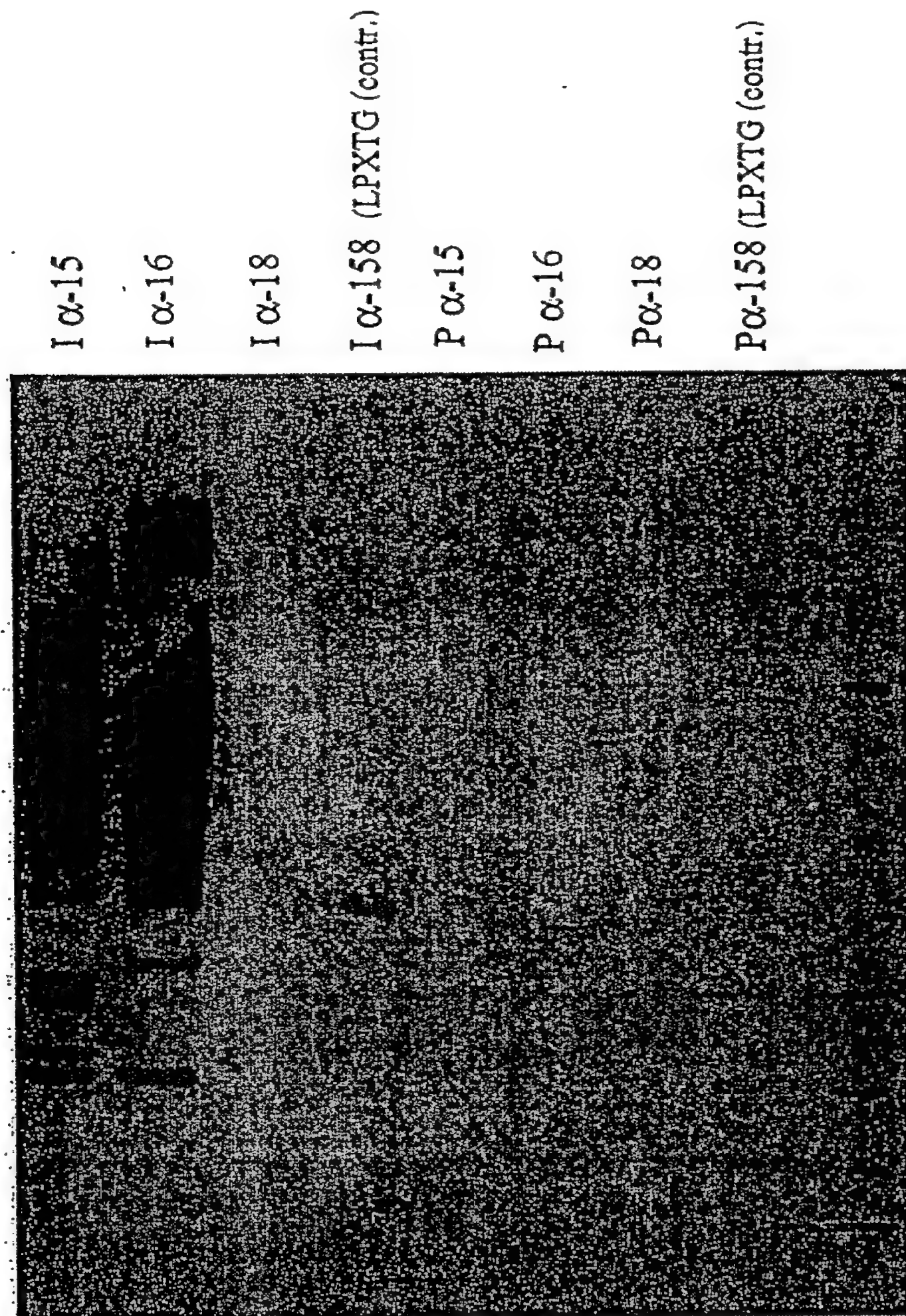


Figure 106

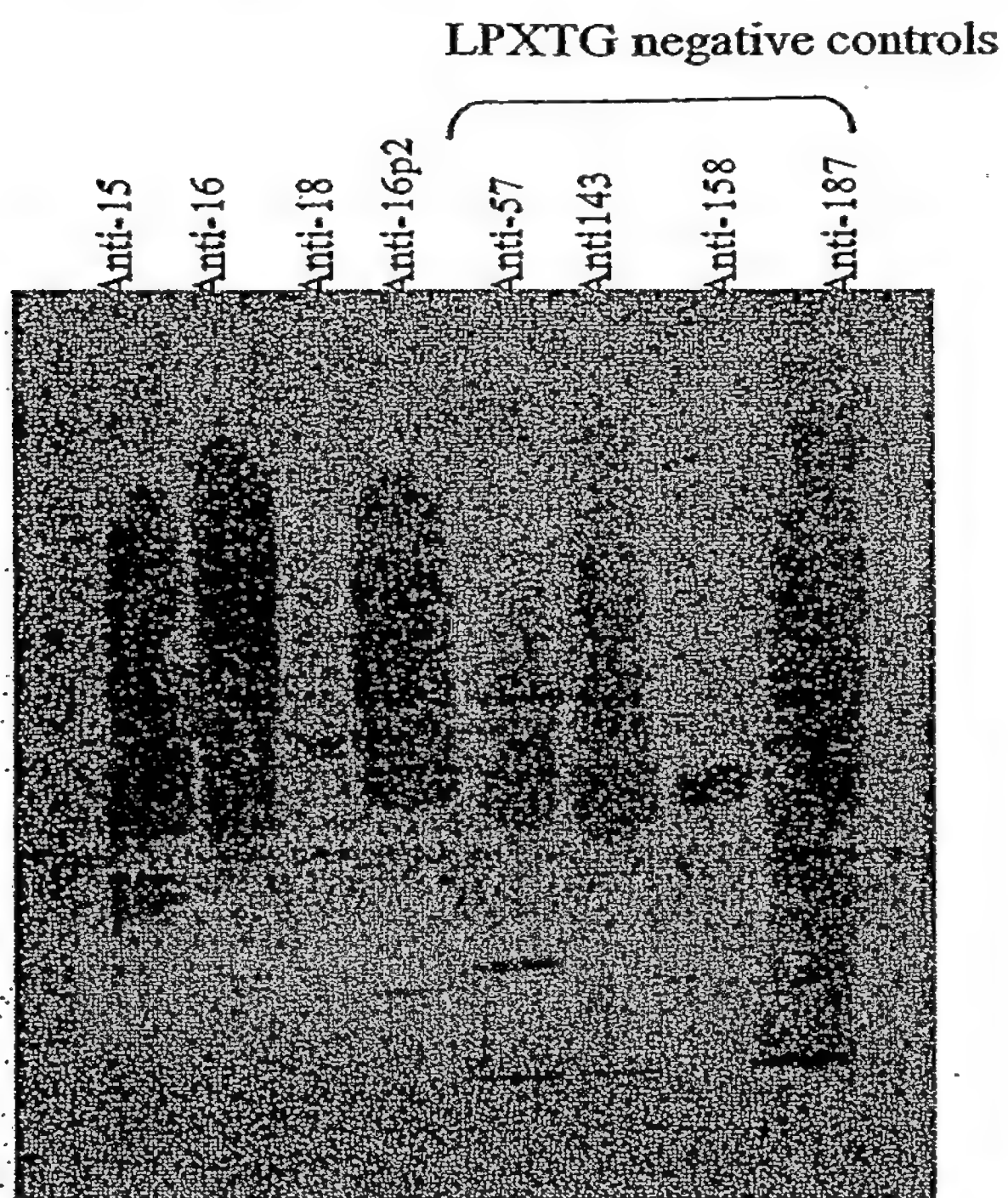
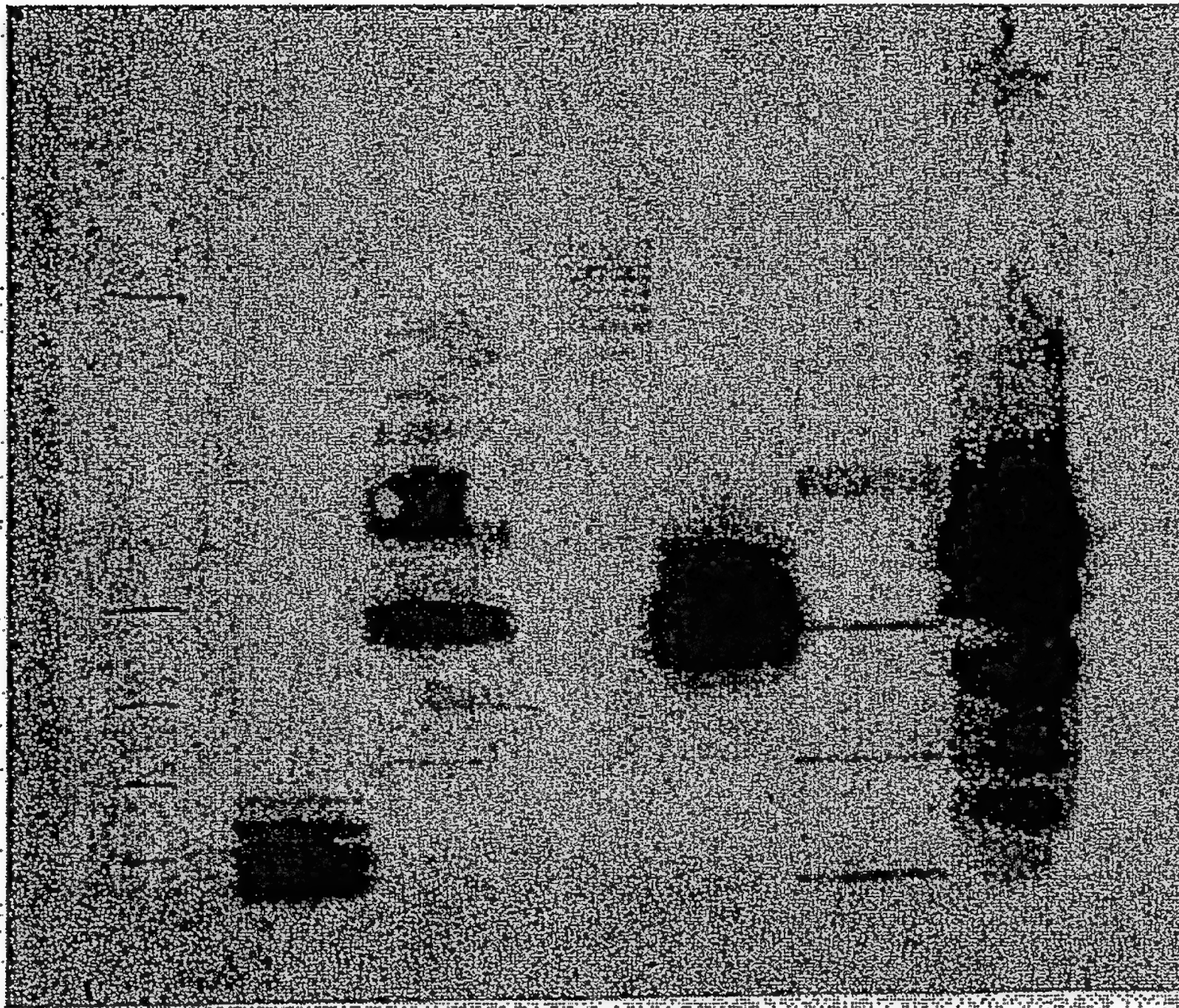
Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)

Figure 107

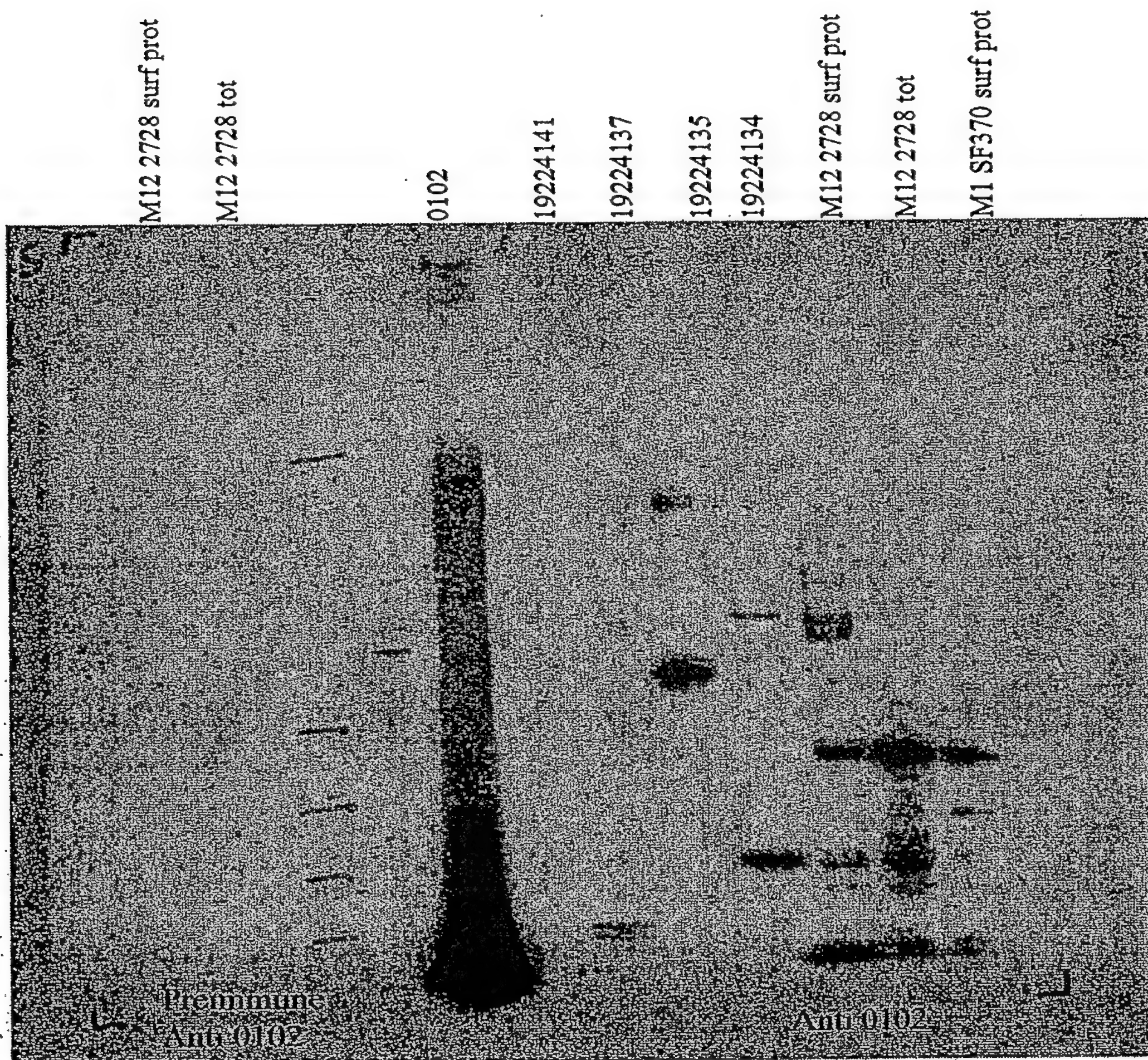
Western blot on fraction enriched in surface proteins of M12 (2728)

Anti-19224134
Anti-19224135
Anti-19224137
Anti-19224141
Anti-0102
Anti-158 (control)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

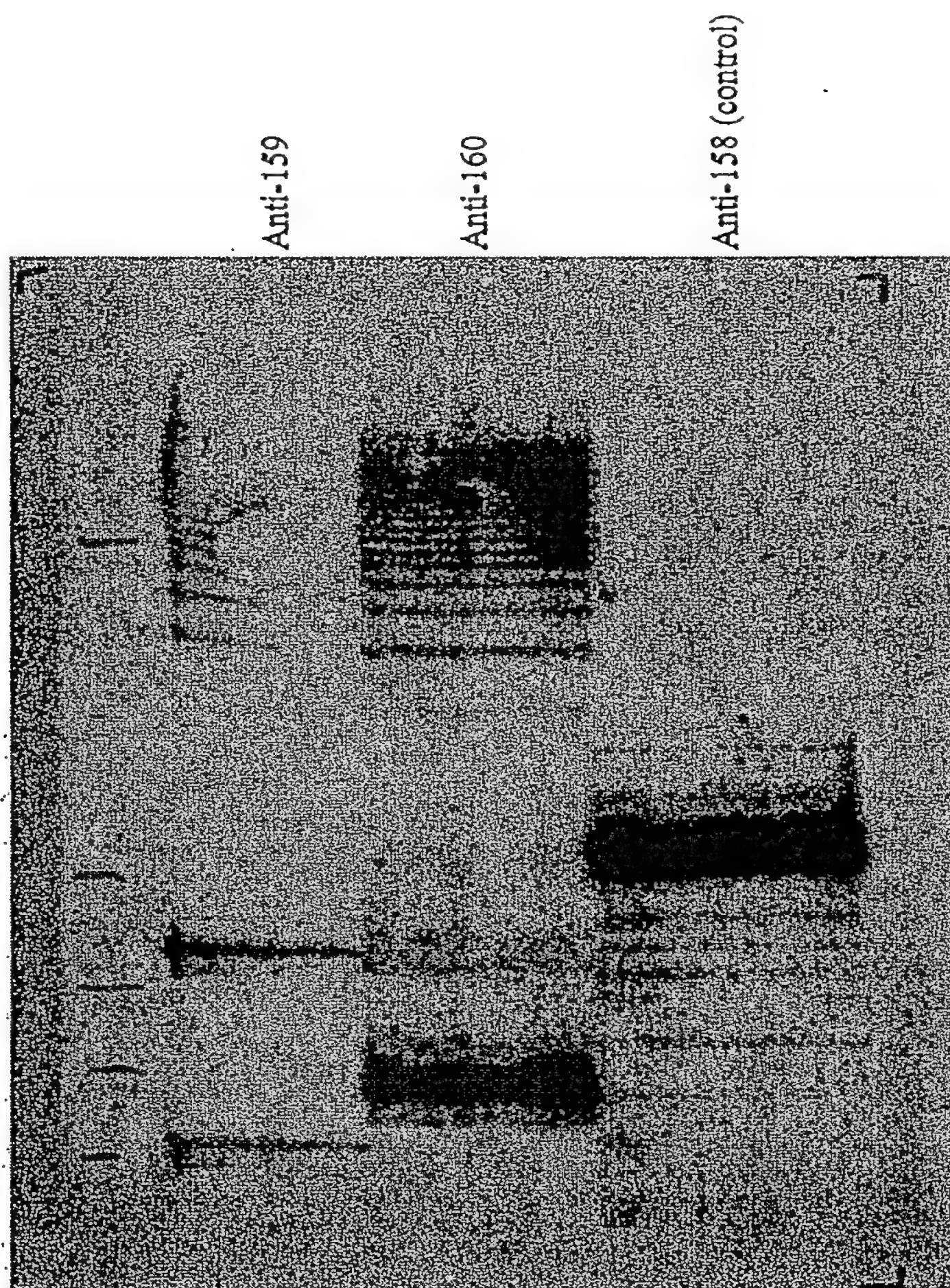
Figure 108



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

Western blot on fraction enriched in surface proteins of M6 (2724)

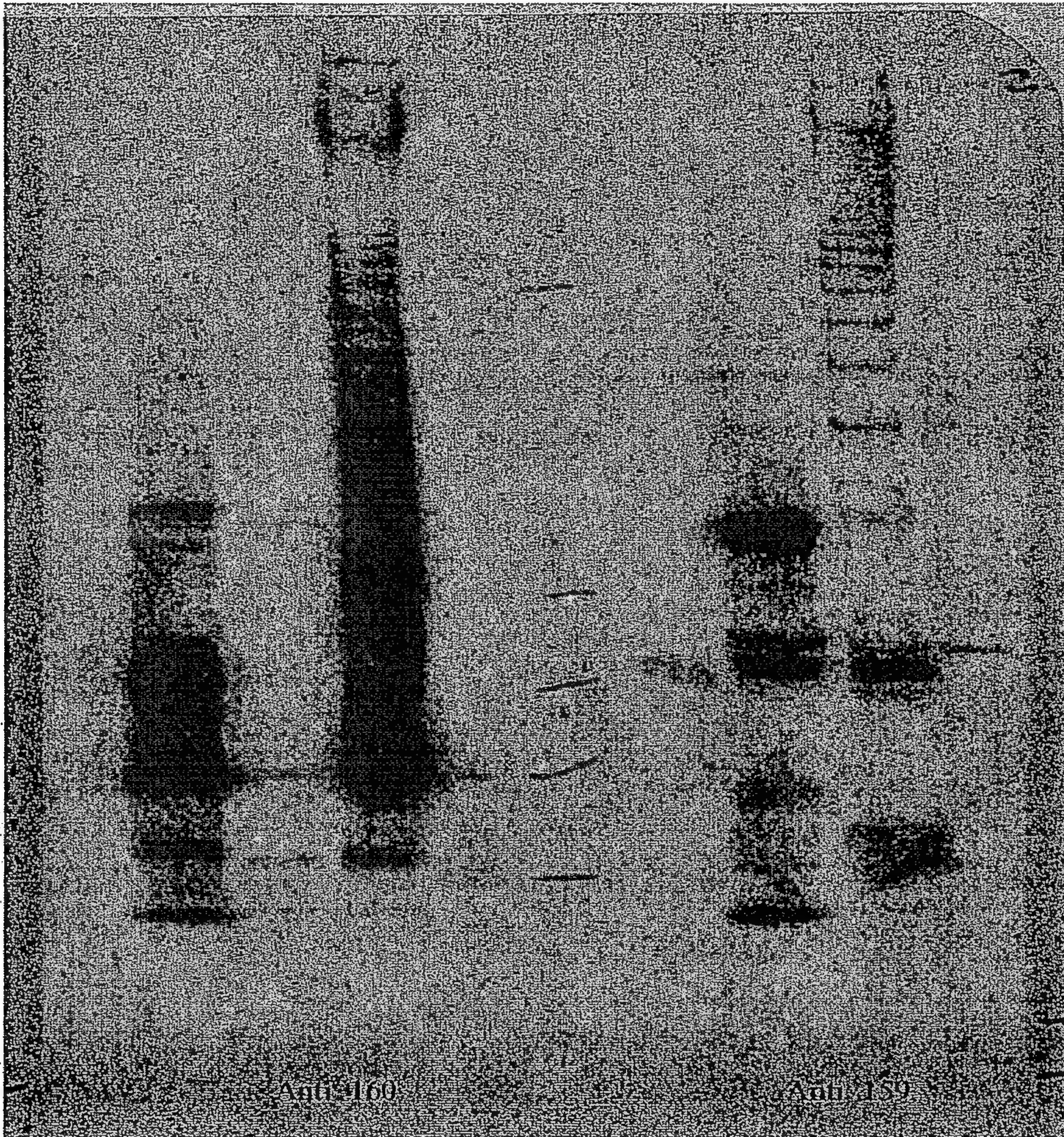


M6 strain isolate 2724

Figure 110

PCT/US05/27239

0160	0159	M6 3650 surf prot	M1 SF370 surf prot	Marker	0160	0159	M6 3650 surf prot	M1 SF370 surf prot
------	------	-------------------	--------------------	--------	------	------	-------------------	--------------------



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

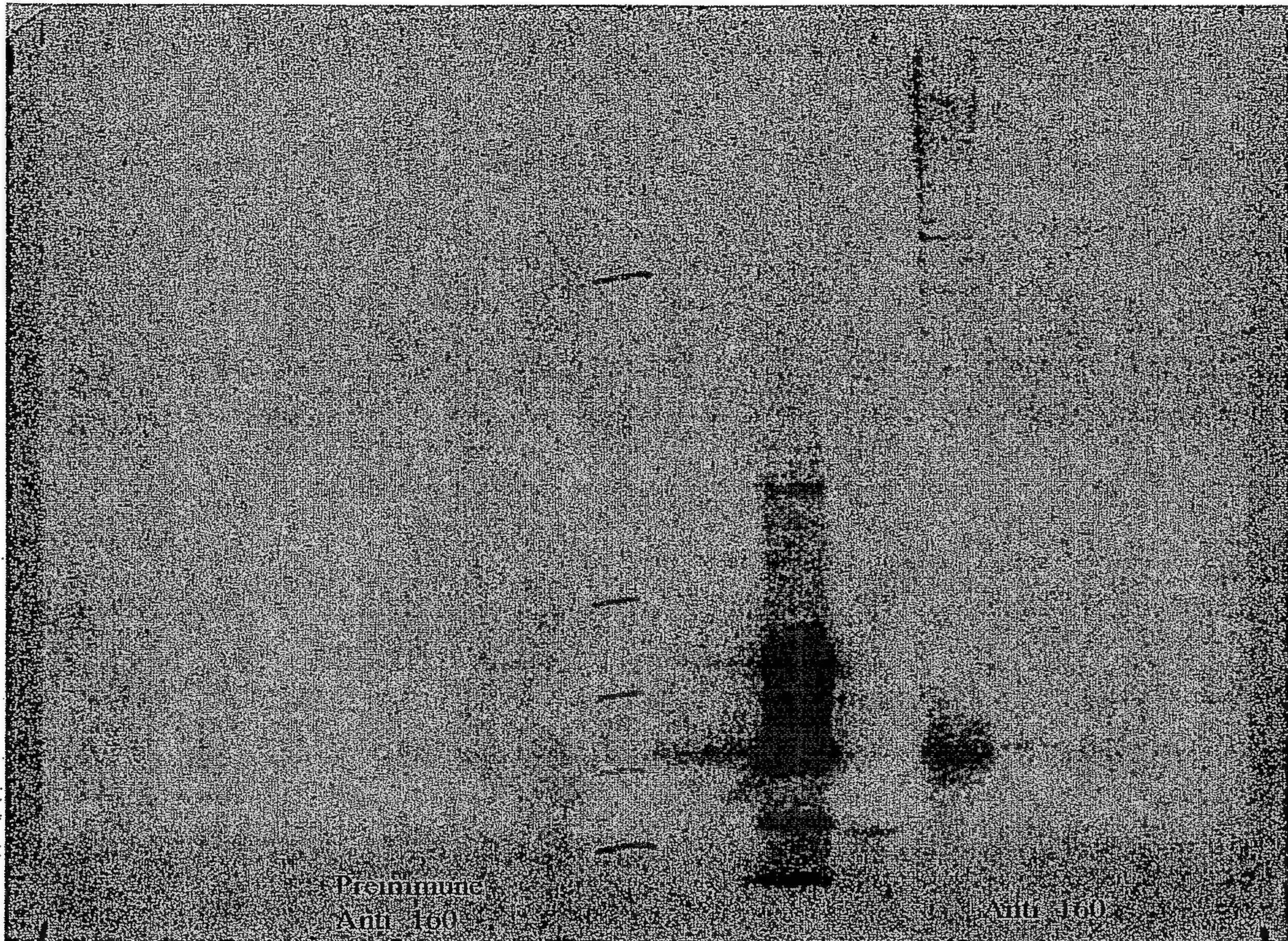


Figure 112

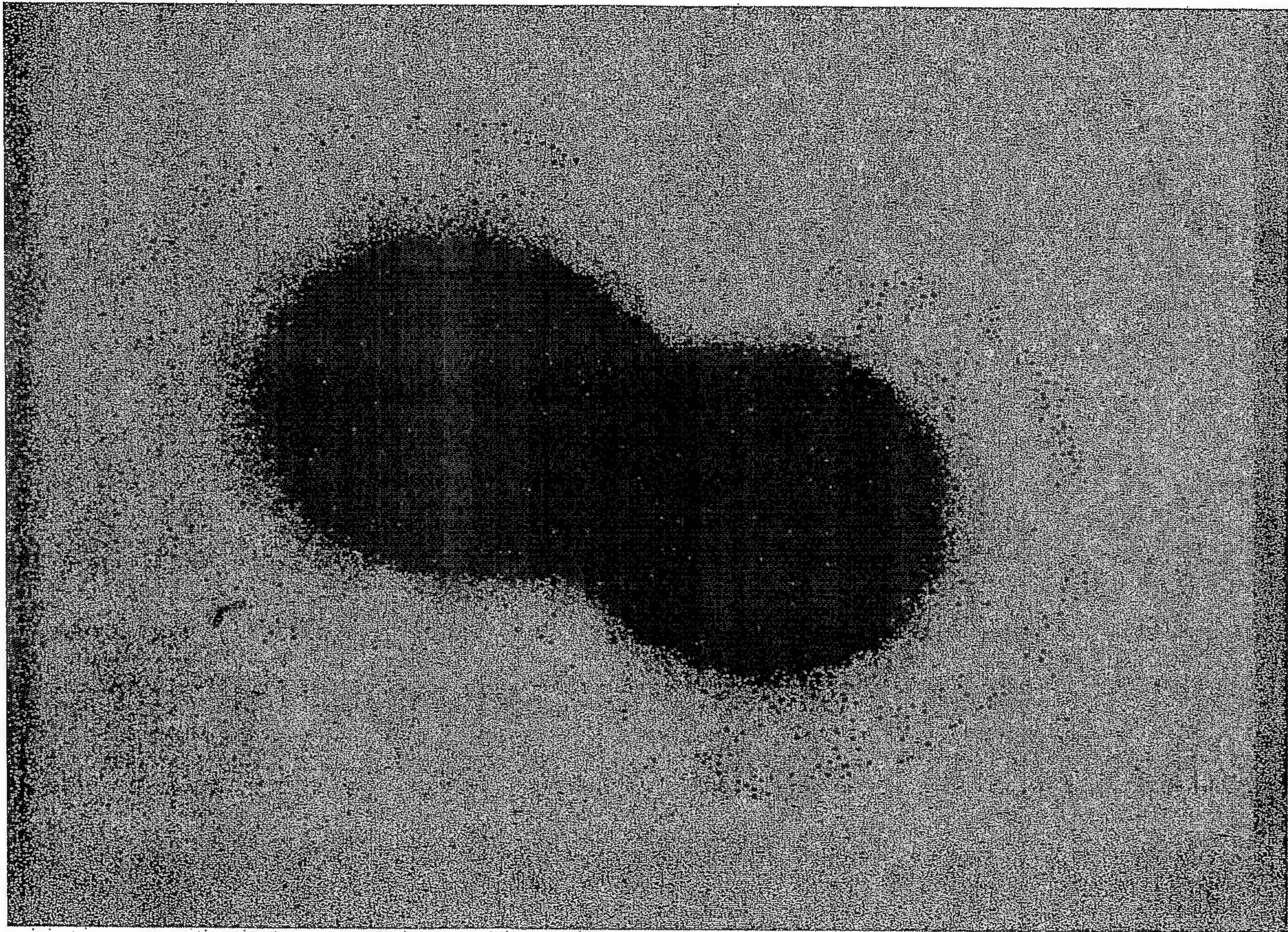


FIGURE 113

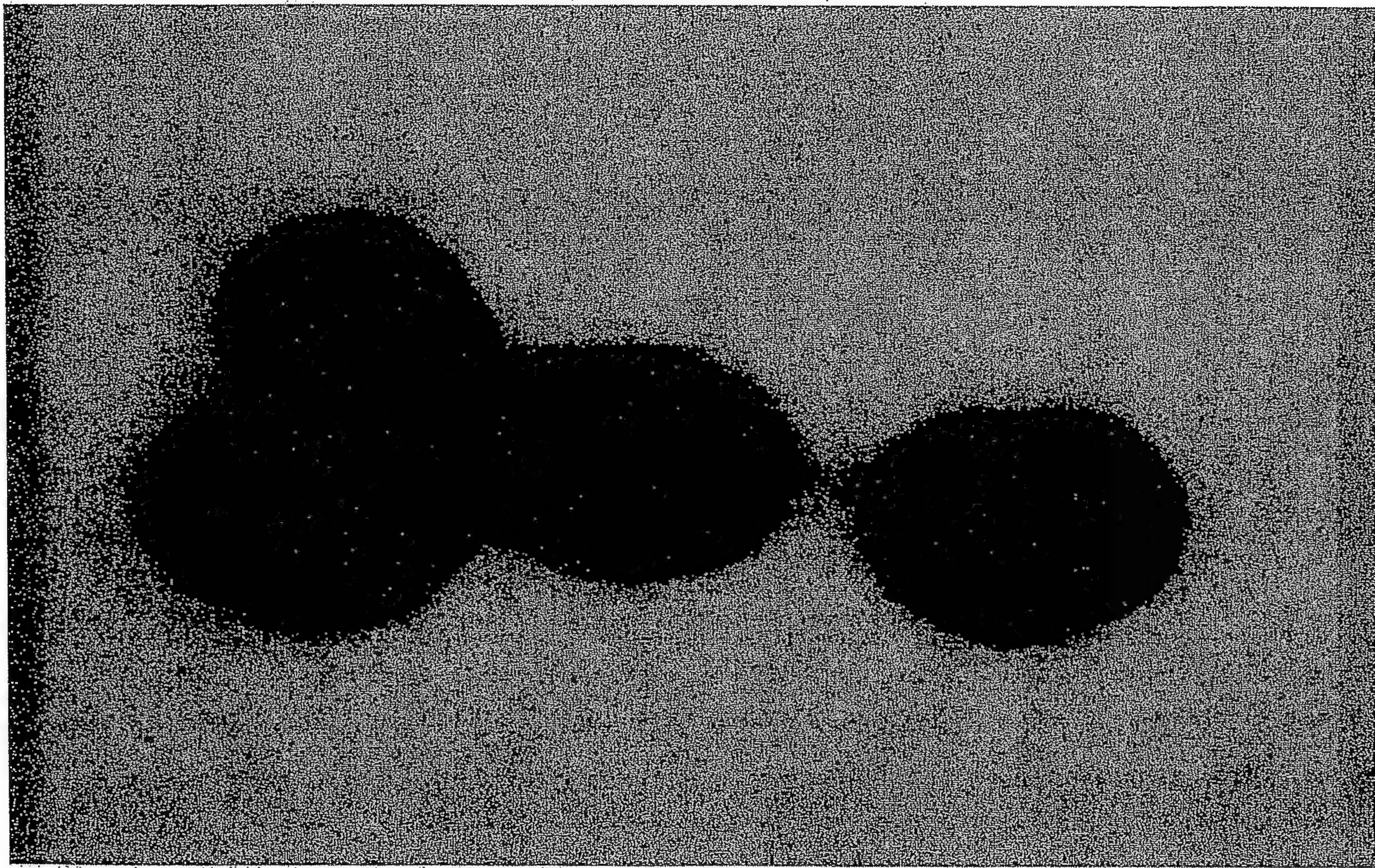


FIGURE 114

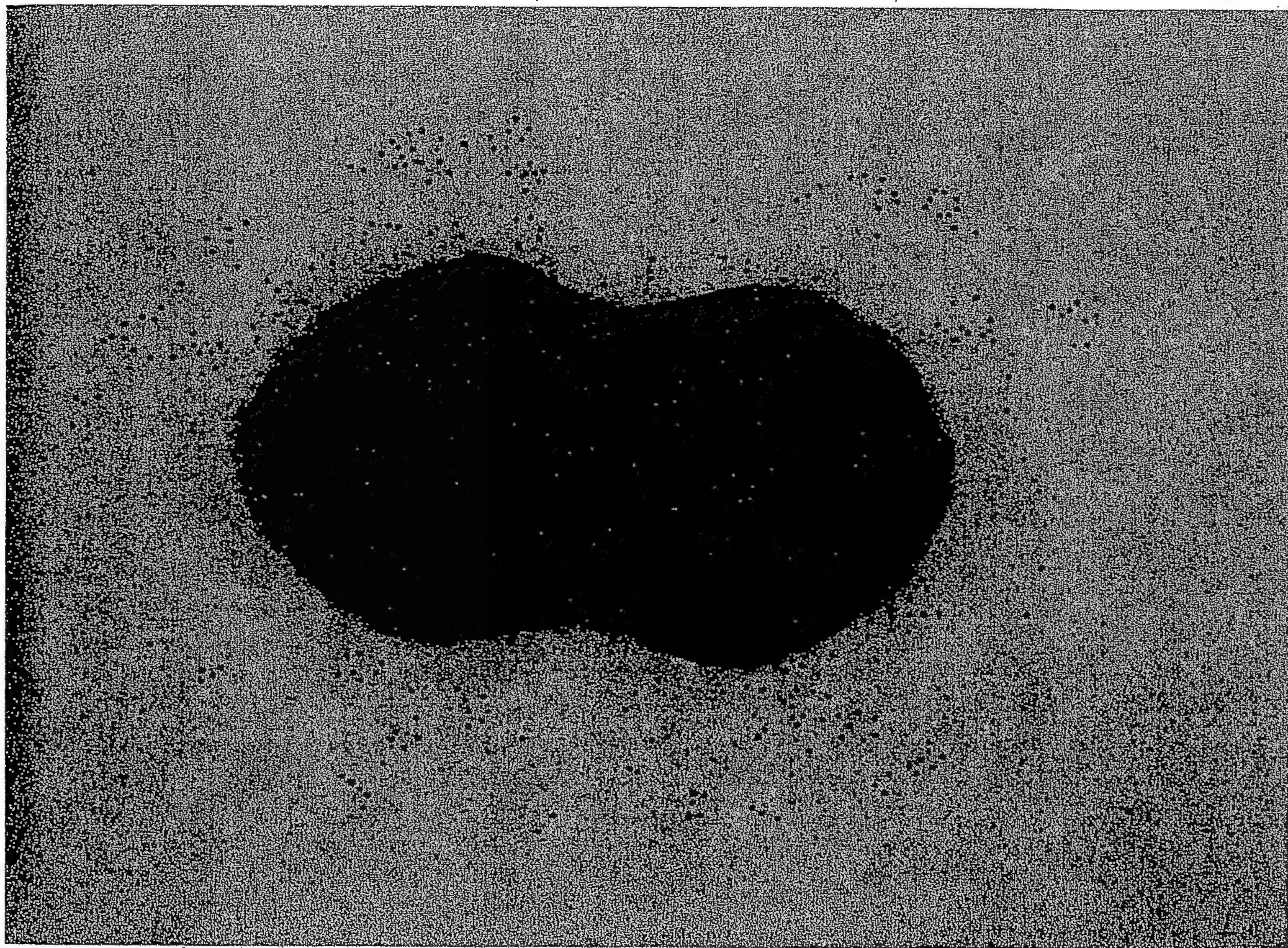


FIGURE 115

PCT/US05/27239 307/487

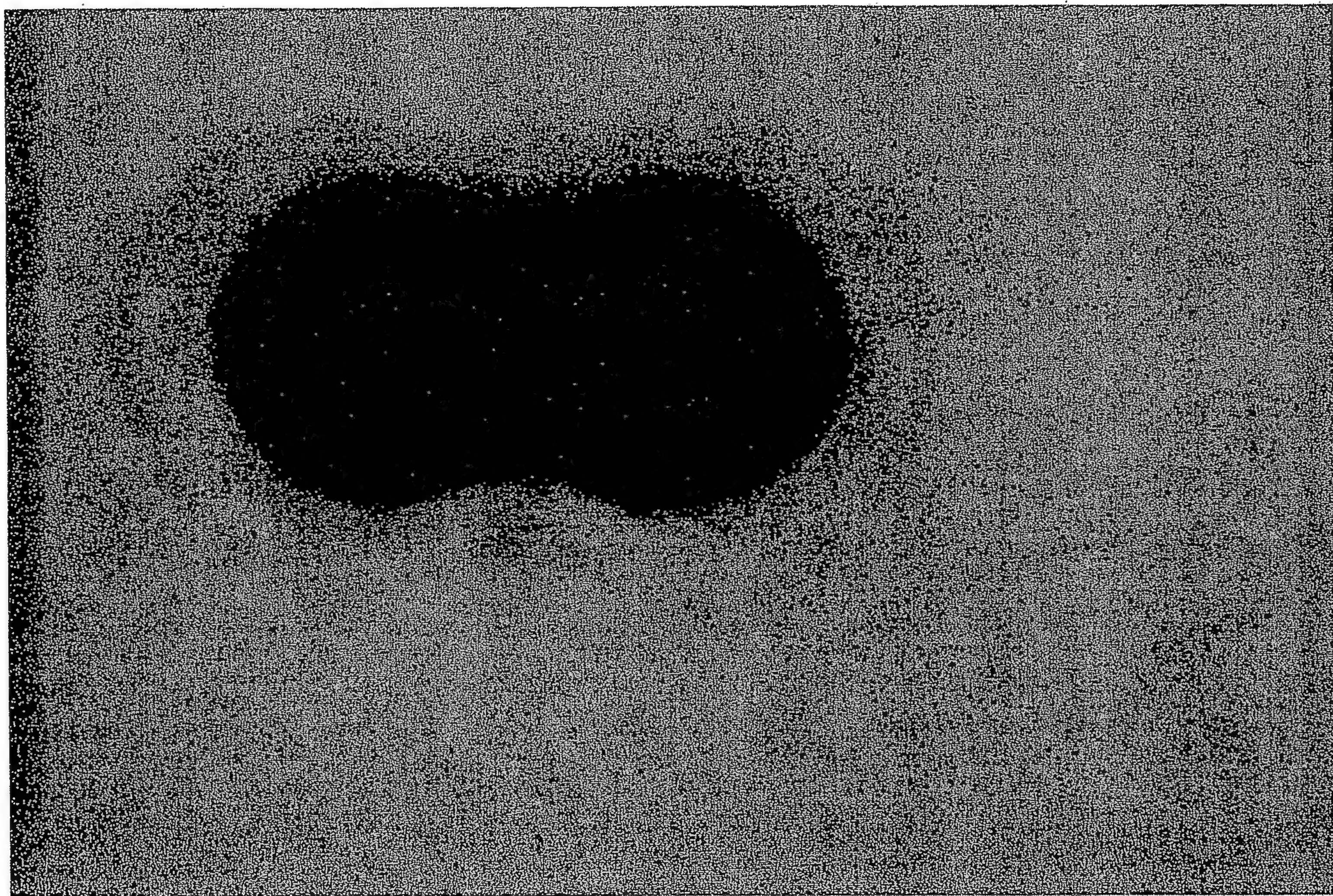


FIGURE 116

PCT/US05/27239

308/487

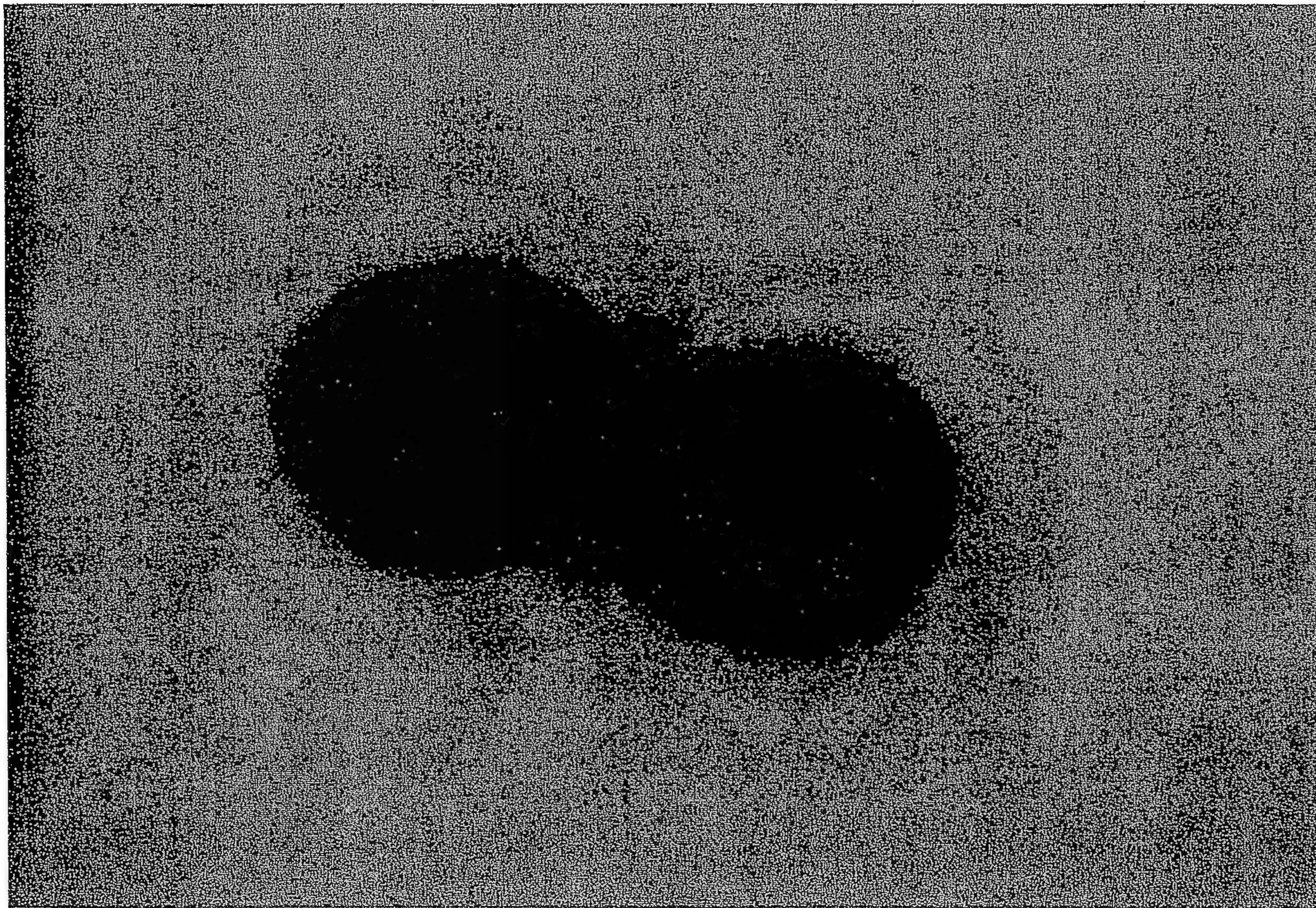


FIGURE 117

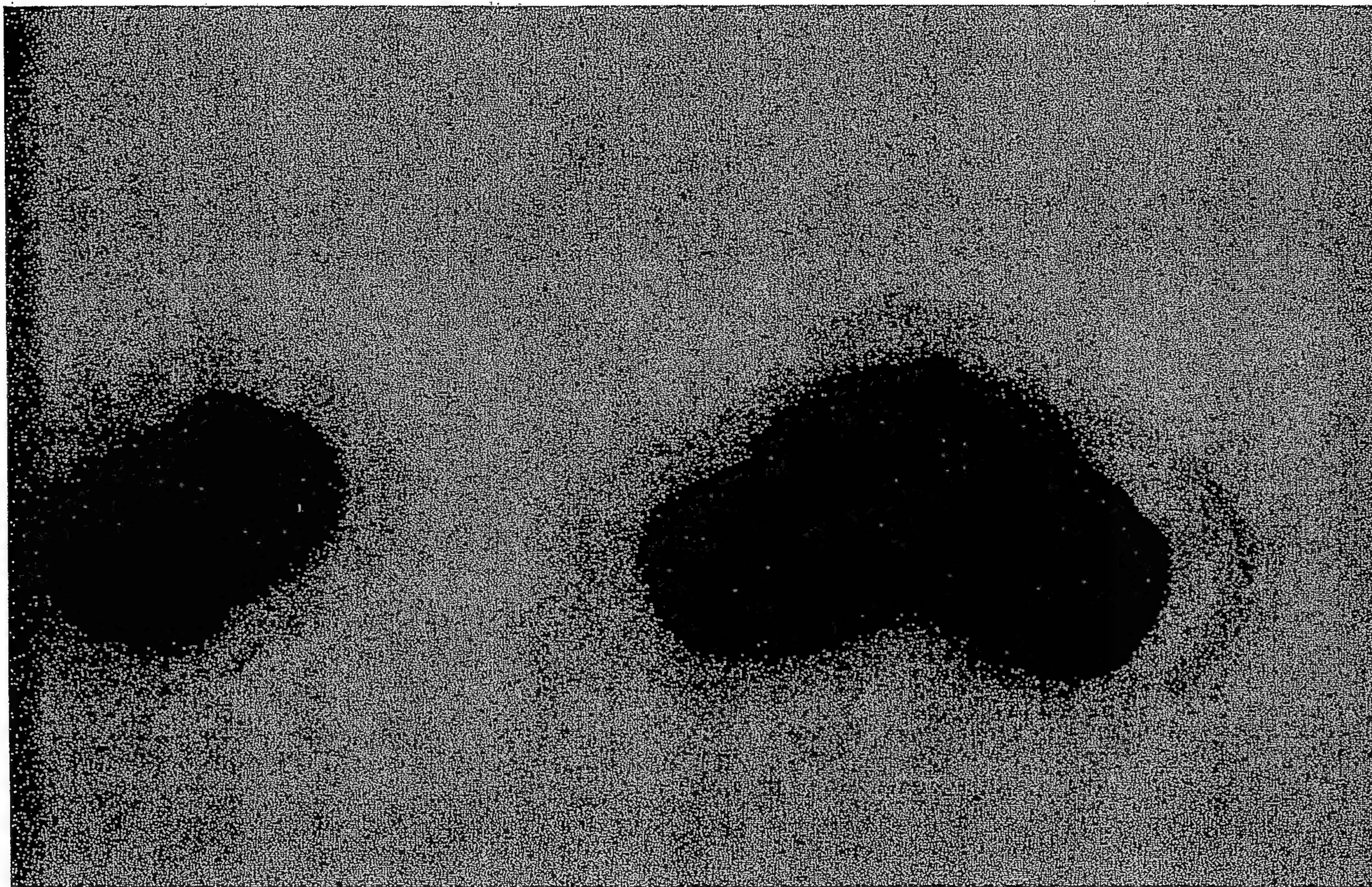


FIGURE 118

PCT/US05/27239

310/487

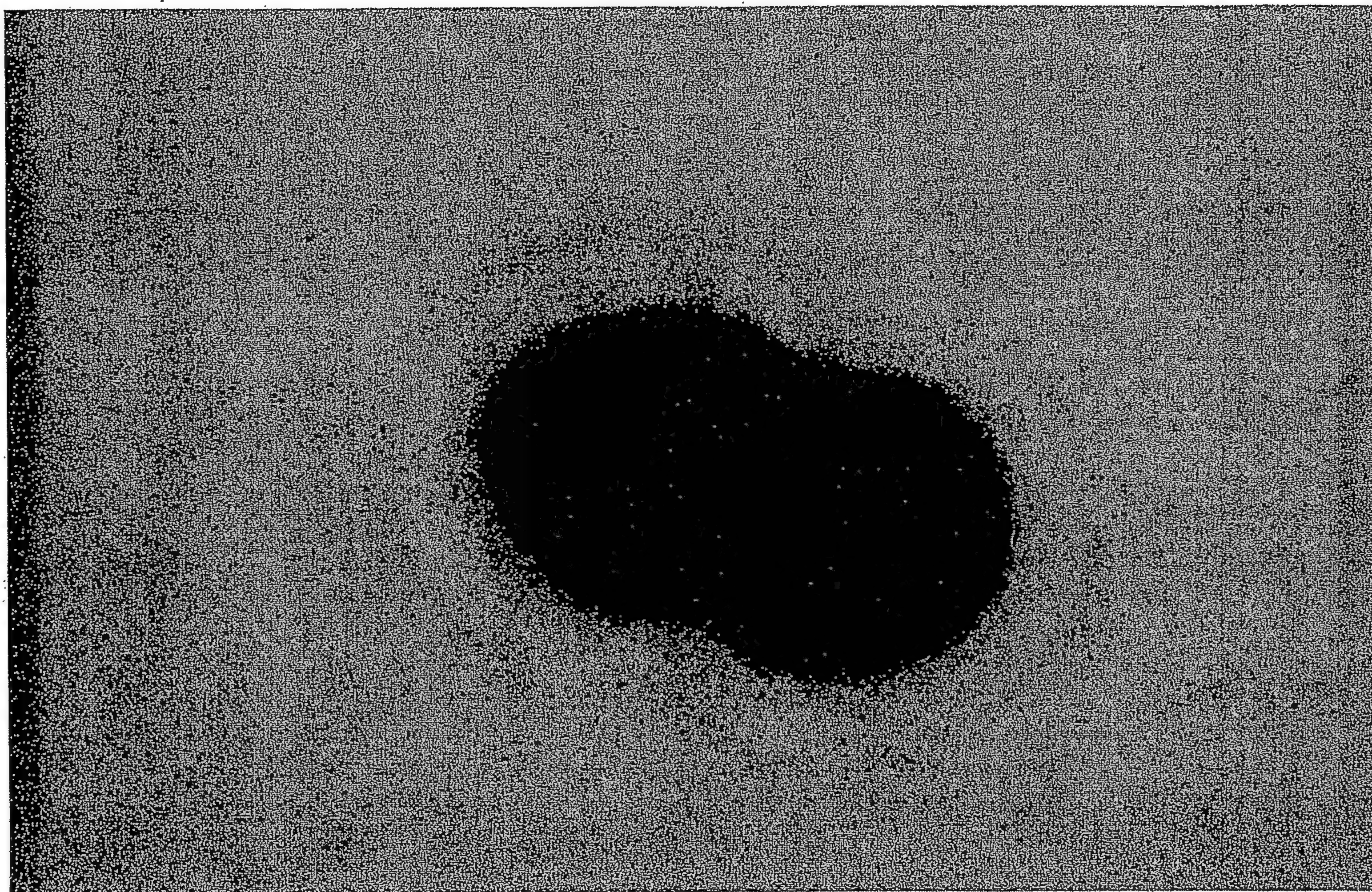


FIGURE 119

PCT/US05/27239

311/487

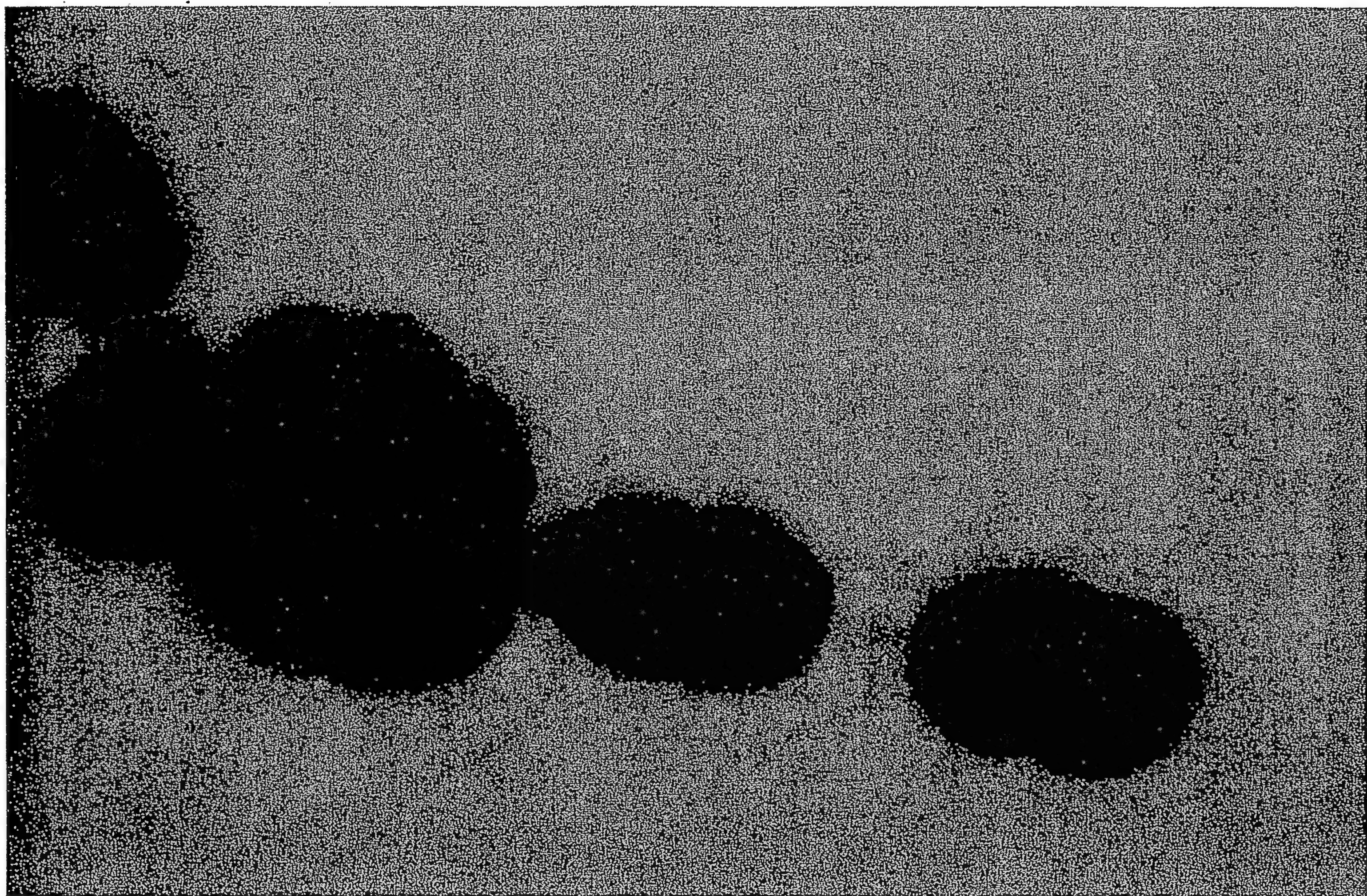


FIGURE 120

PCT/US05/27239

312/487

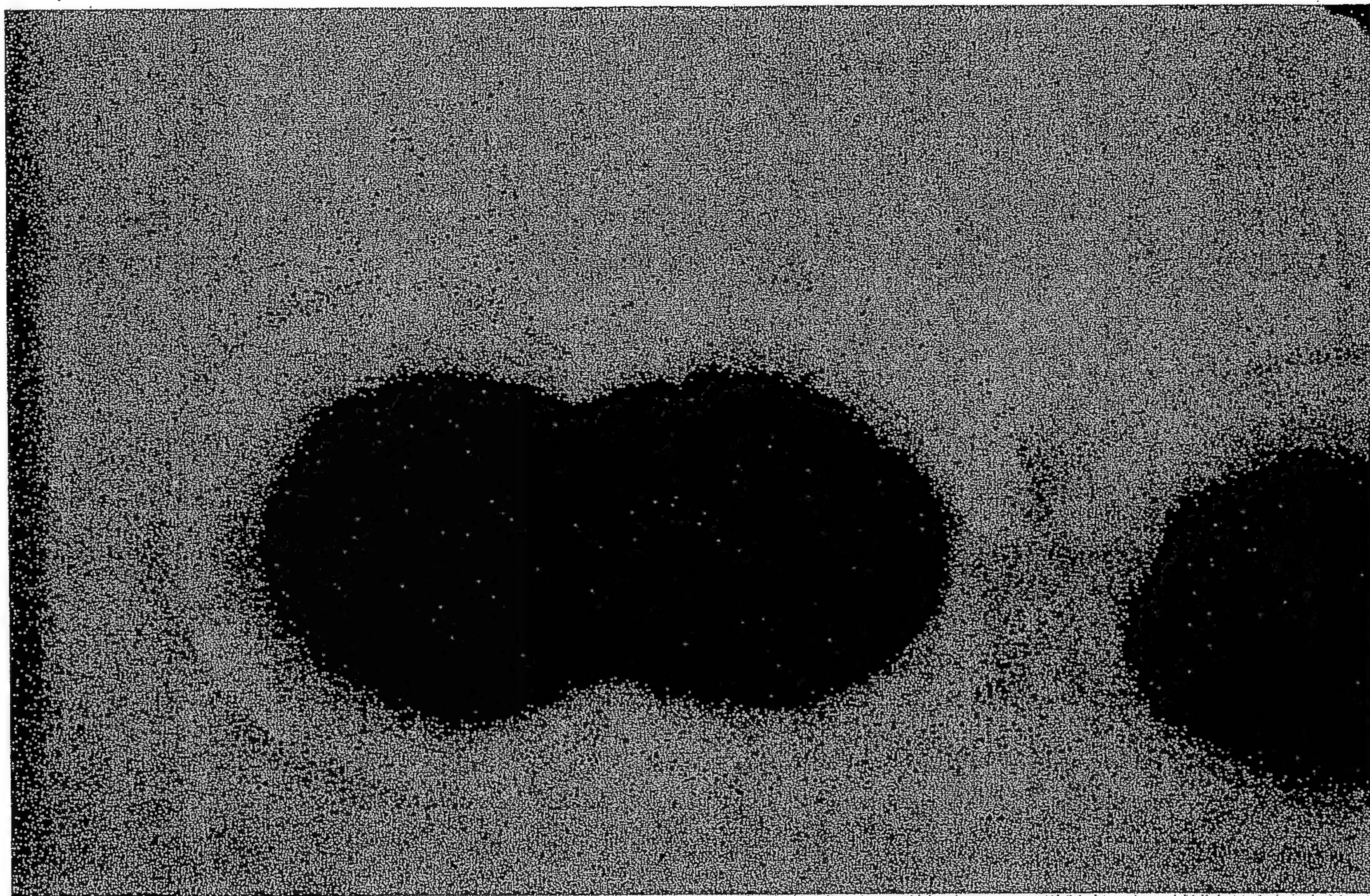


FIGURE 121

PCT/US05/27239

313/487

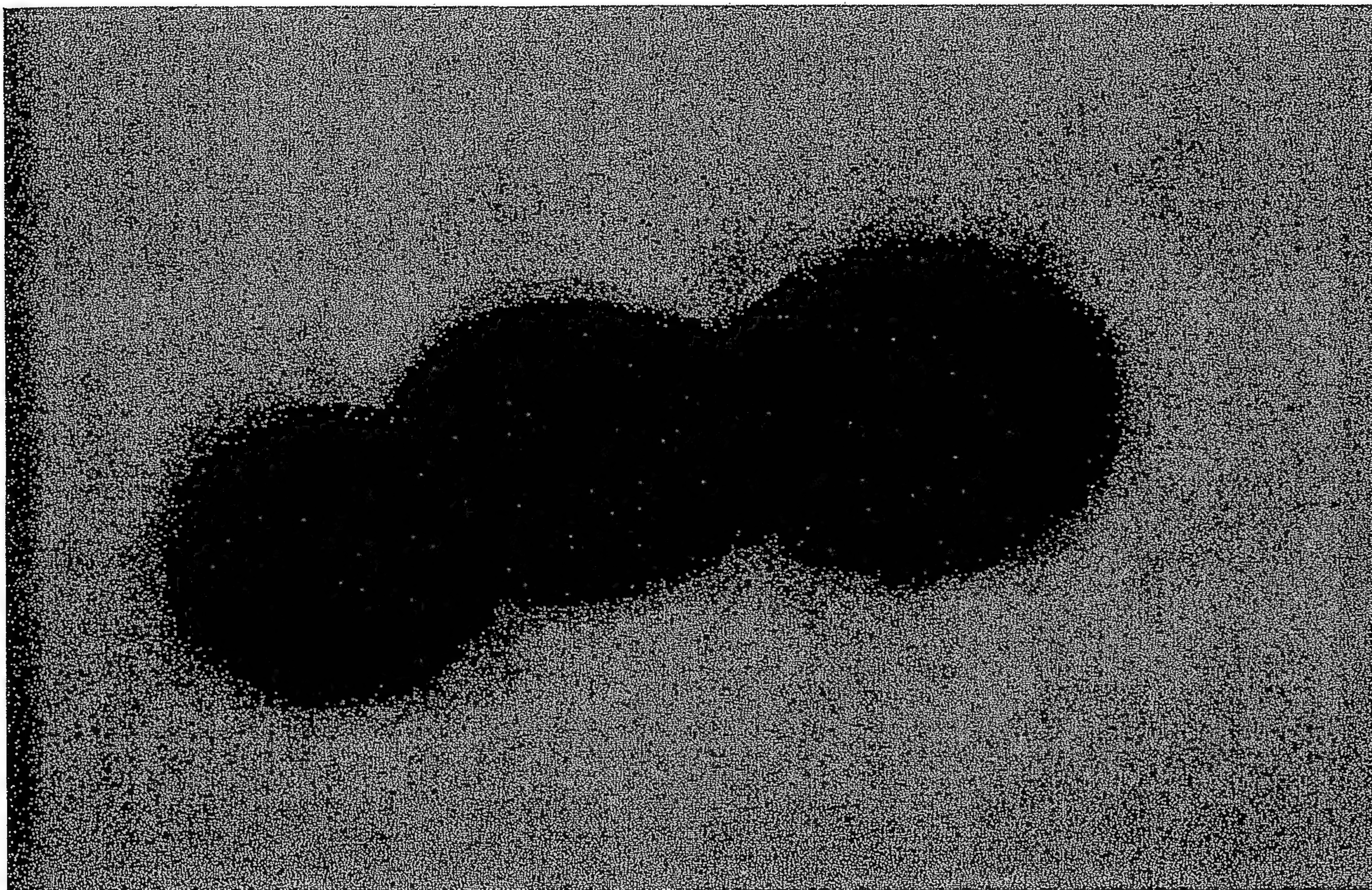


FIGURE 122

PCT/US05/27239

314/487

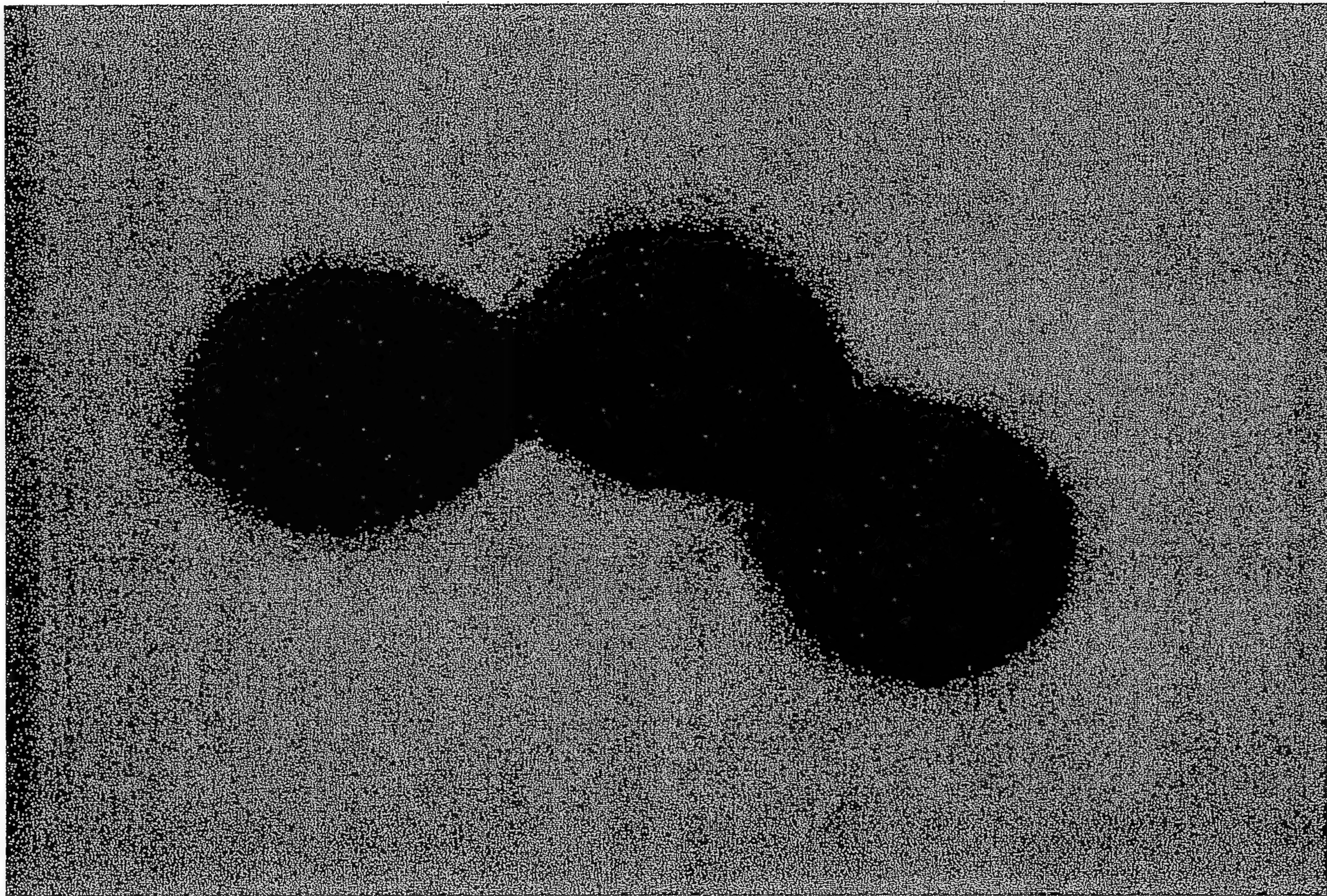


FIGURE 123

315/487

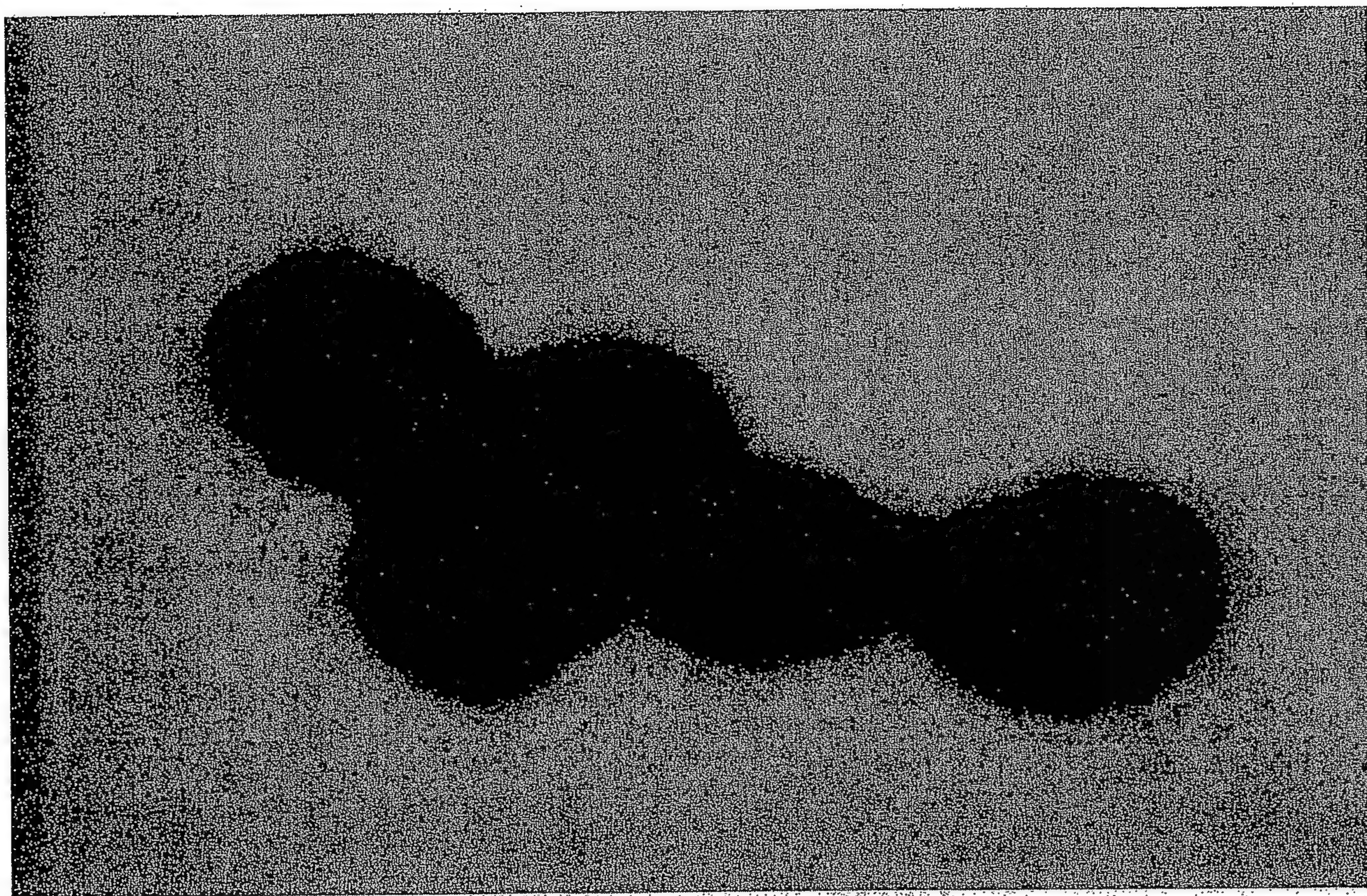


FIGURE 124

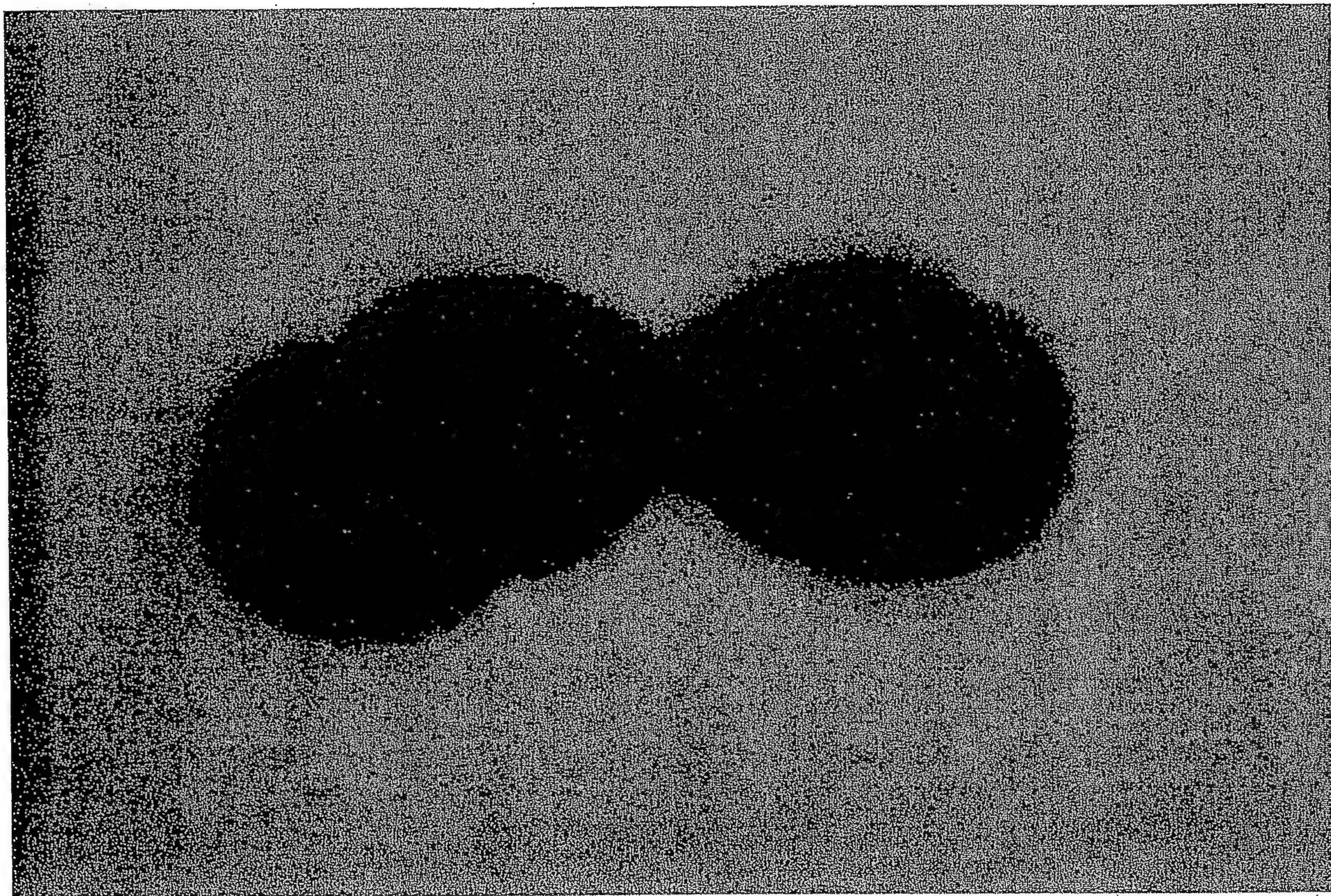


FIGURE 125

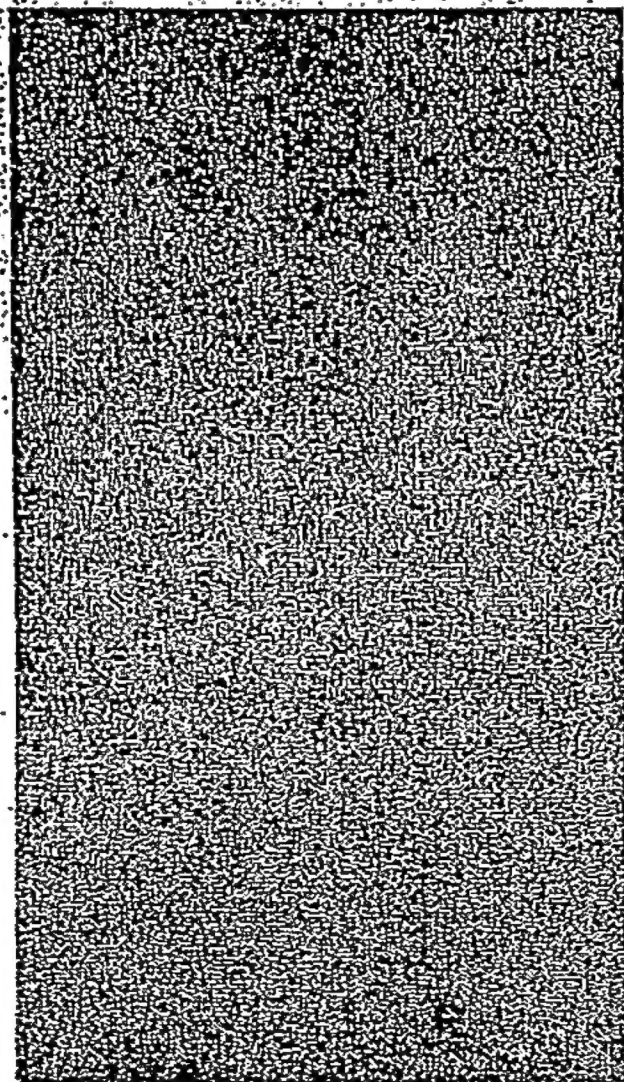
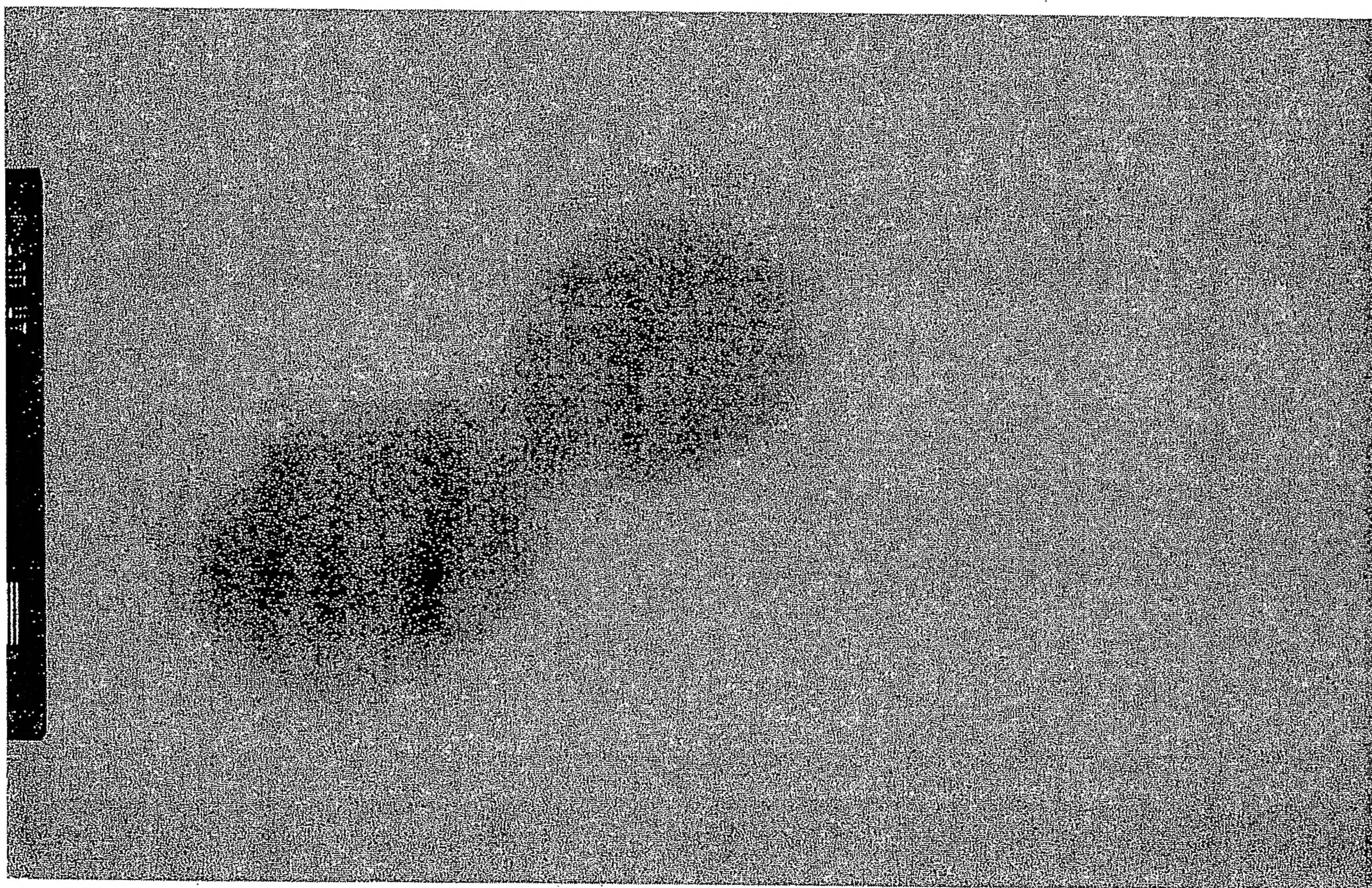


FIGURE 126

Figure 127



PCT/US05/27239 319/487

Figure 128

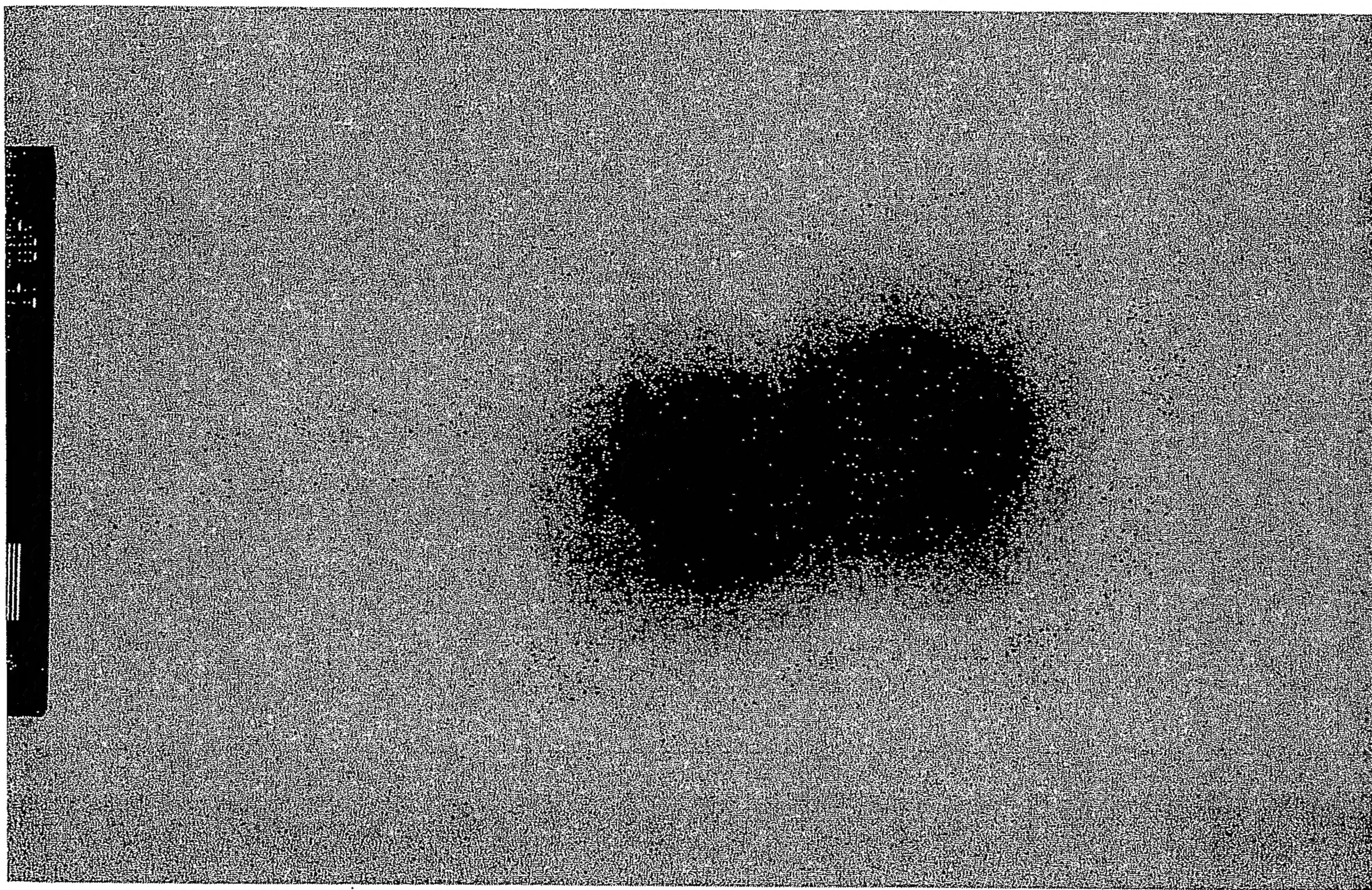


Figure 129

